

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 ; Search time 23.93 Seconds

(without alignments)
151.675 Million cell updates/sec

Title: US-09-582-711-3

Perfect score: 252

Sequence: 1 STGHSGSHSHSTTTGGRSDA.....TSRETRDQSGSGSRHSGS 49

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.GeneSeq_1101.*
1: /SIDS8/gcgcdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgcdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgcdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgcdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgcdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgcdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgcdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgcdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgcdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgcdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgcdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgcdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgcdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgcdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgcdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgcdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgcdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgcdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgcdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgcdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgcdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgcdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No | Score | Query Match | Length | ID | Description |
|-----------|-------|-------------|--------|----------|---------------------|
| 1 | 252 | 100.0 | 49 | AAW50747 | Peptide S-47-S derl |
| 2 | 252 | 100.0 | 49 | AAW23931 | Flaggrin-derived |
| 3 | 208 | 82.5 | 330 | AAW22954 | Human flaggrin seq |
| 4 | 208 | 82.5 | 330 | AAW22955 | Human flaggrin seq |
| 5 | 208 | 82.5 | 330 | AAW22956 | Human flaggrin seq |
| 6 | 207 | 82.1 | 330 | AAW22957 | Human flaggrin seq |
| 7 | 73.5 | 29.2 | 982 | AAH13320 | Murine Natural Kll |
| 8 | 72.5 | 28.8 | 63 | AAW21759 | Peptide #8193 enco |
| 9 | 70.5 | 28.0 | 407 | AAW38081 | Peptide #12118 enc |
| 10 | 70.5 | 28.0 | 1275 | AAW29084 | Human protein sequ |
| 11 | 70.5 | 28.0 | 20 | AAW29084 | Human C1F130 prote |

| | | | | | | |
|----|------|------|------|----|----------|--------------------|
| 12 | 70.5 | 28.0 | 1275 | 21 | AAW44262 | Human cell cycle r |
| 13 | 69 | 27.4 | 717 | 22 | AAW31581 | Amino acid sequenc |
| 14 | 67 | 26.6 | 306 | 21 | AAW54149 | Human pancreatic c |
| 15 | 67 | 26.6 | 443 | 22 | AAW40059 | Human polypeptide |
| 16 | 67 | 26.6 | 574 | 22 | AAW41855 | Human polypeptide |
| 17 | 66.5 | 26.4 | 330 | 21 | AAW70227 | Human RNA-associat |
| 18 | 66.5 | 26.4 | 594 | 22 | AAW12176 | Human PRO498 poly |
| 19 | 66.5 | 26.4 | 674 | 21 | AAW53044 | Human secreted pro |
| 20 | 66 | 26.2 | 860 | 19 | AAW63700 | Receptor type tyro |
| 21 | 65 | 25.8 | 413 | 22 | AAW72163 | Human RNA metaboli |
| 22 | 65 | 25.8 | 655 | 22 | AAW70771 | Murine neutral cer |
| 23 | 64 | 25.4 | 16 | 20 | AAW22952 | Synthetic peptide |
| 24 | 64 | 25.4 | 3025 | 10 | AAW3284 | Sequence of clone |
| 25 | 63.5 | 25.2 | 280 | 21 | AAW24816 | Arabidopsis thalia |
| 26 | 63.5 | 25.2 | 282 | 21 | AAW64630 | Arabidopsis thalia |
| 27 | 63.5 | 25.2 | 288 | 21 | AAW24815 | Arabidopsis thalia |
| 28 | 63.5 | 25.2 | 290 | 21 | AAW64629 | Arabidopsis thalia |
| 29 | 63.5 | 25.2 | 326 | 21 | AAW24814 | Arabidopsis thalia |
| 30 | 63.5 | 25.2 | 328 | 21 | AAW64628 | Arabidopsis thalia |
| 31 | 63 | 25.0 | 529 | 19 | AAW80498 | A protein designat |
| 32 | 63 | 25.0 | 529 | 19 | AAW85055 | Human protein desl |
| 33 | 63 | 25.0 | 529 | 21 | AAW44989 | Human epidermal pr |
| 34 | 63 | 25.0 | 759 | 22 | AAW94613 | Human protein sequ |
| 35 | 63 | 25.0 | 1850 | 22 | AAW72210 | Modified chicken v |
| 36 | 62.5 | 24.8 | 189 | 22 | AAW82799 | S. epidermidis ope |
| 37 | 62 | 24.6 | 61 | 22 | AAW18795 | Peptide #5329 enco |
| 38 | 62 | 24.6 | 61 | 22 | AAW31309 | Peptide #5346 enco |
| 39 | 62 | 24.6 | 98 | 21 | AAW01314 | Human secreted pro |
| 40 | 62 | 24.6 | 234 | 22 | AAW94508 | Human protein sequ |
| 41 | 62 | 24.6 | 276 | 21 | AAW56814 | Human prostate can |
| 42 | 62 | 24.6 | 387 | 22 | AAW62626 | Human RNA helicase |
| 43 | 62 | 24.6 | 431 | 22 | AAW44995 | Human colon cancer |
| 44 | 62 | 24.6 | 442 | 22 | AAW40816 | Human polypeptide |
| 45 | 62 | 24.6 | 476 | 19 | AAW57241 | Heterodera glycine |

ALIGNMENTS

RESULT 1
AAW50747 standard; Protein: 49 AA.
ID AAW50747:
AC AAW50747:
XX
XX 31-JUL-1998 (first entry)
XX
XX Peptide S-47-S derived from flaggrin.
DE
XX
XX Flaggrin: antigen; anti-flaggrin; rheumatoid polyarthritiis;
KW
KW citrulline; antibody.
XX
XX Synthetic.
OS Homo sapiens.
XX
XX W09808946-A1.
XX
XX 05-MAR-1998.
XX
XX 01-SEP-1997; 97W0-FR01541.
XX
XX 30-AUG-1996; 96FR-0010651.
XX
XX (INMR) BIOMERIEUX SA.
XX
XX Arnaud M, Dalbon P, Jolivet M;
PI Sebbaq M, Serre G, Simon M, Vincent C;
PI Girbal-Neuhausser E, Jolivet-Reynaud C;
XX
XX WPI: 1998-207042/18.
XX
XX Artificial antigen recognised by anti-flaggrin auto-antibodies - 1s
PT modified form of flaggrin with citrulline replacing at least one

| | |
|----------|--|
| OS | Homo sapiens. |
| XX | |
| PN | MO9928344-A2. |
| XX | |
| PD | 10-JUN-1999. |
| XX | |
| PF | 30-NOV-1998; 98MO-EP07714. |
| XX | |
| PR | 09-APR-1998; 98EP-0870078. |
| XX | |
| PR | 28-NOV-1997; 97EP-0870195. |
| XX | |
| PA | (INNO-) INNOGENETICS NV. |
| XX | |
| PI | Meheus L, Raymackers J, Union A; |
| XX | |
| DR | WPI; 1999-385357/32. |
| XX | |
| PT | New peptide derived from intermediate filament proteins |
| XX | |
| PS | Example 1; Fig 2; 73pp; English. |
| XX | |
| CC | AA22954-57 represent amino acid sequences of human filaggrin clones. The |
| CC | specificity of these peptides is described. The peptides are derived from any variant of natural |
| CC | filaggrin or any variant of intermediate filament proteins. These |
| CC | peptides contain at least one citrulline residue which is crucial |
| CC | for reacting with antibodies that are present in sera from patients |
| CC | with rheumatoid arthritis. The peptides constitute immunogenic |
| CC | determinants of antibodies present in patients with rheumatoid |
| CC | arthritis. The peptides, antibodies, immunotoxins and intermediate |
| CC | filament proteins can be used for the preparation of a therapeutic or |
| CC | of a diagnostic for rheumatoid arthritis. The peptides can also be |
| CC | used for identifying compounds which modulate the interaction between |
| CC | an autoantigen and a rheumatoid arthritis specific autoantibody. The |
| CC | products can also be used for the diagnosis and treatment of other |
| CC | autoimmune diseases e.g. systemic lupus erythematosus, discoid lupus |
| CC | erythematosus, scleroderma, dermatomyositis, or Sjogrens syndrome. |
| XX | |
| Sequence | 330 AA; |

PF 08-JUN-1990; 90US-0143578.
XX
PR 08-JUN-1990; 90US-0535206.
XX
PA (USSH) NAT INST OF HEALTH.
XX
PI Ortaldo J, Young H, Anderson S;
XX
DR WPI, 1991-245694/33.
DR N-PSDB; AAQ13115.
XX
PT DNA encoding a natural killer cell receptor - used to develop
PT prods. for the immuno-detection and immuno-therapy of tumours
XX
PS Disclosure; Fig 2; 30pp; English.
XX
XX Overlapping clones, which make up the cDNA sequence from which this
XX sequence was deduced, were isolated from a murine PBL cDNA library
CC prepared in lambda gt10. The purified protein can mediate the
CC cytolytic activity of mammalian cells. It specifically distinguishes
CC tumour cells making it a candidate for the development of products
CC for the immunodetection and immunotherapy of tumours.
CC See also AAQ13114.
XX
XX
SQ Sequence 982 AA;

| | | | | | | | |
|-----------------------|-------|---|-------|----------------------------------|------|-----------|-----|
| Query Match | 29.2% | Score 73.5 | DB 12 | length 982 | | | |
| Best Local Similarity | 39.3% | Pred. No. 1.5 | | | | | |
| Matches | 22 | Conservative | 6 | Mismatches 21; Indels 7; Gaps 2; | | | |
| OY | 1 | STGAGCGSHMTTTCGRS | --- | DASGSSGSRKTSRSTETRDQDS | ---- | GDGSHRSGS | 49 |
| Db | 693 | ssgagagsssrssrsrsrstrsrslptsrslpsrstrsrshpksydcgaghsr | | ssr | | ssr | 748 |

| | | | | |
|-----------------------|-------|---|--------|-------------------------|
| Query Match | 82.1% | Score 207; | DB 20; | Length 330; |
| Best Local Similarity | 81.2% | Pred. No. 5, 1e-17; | | |
| Matches | 39; | Conservative | 4; | Mismatches 5; Indels 0; |
| OY | 2 | TGHSSTSHHTTQGRSDASRSGSSRSTSRRTROEGSGDGSRRSGS | 49 | |
| | | | | |
| Db | 75 | sgnshstshhtcstsggrsdasnghsrstsrstsrtrndesgddsrstsgs | 122 | |

| | | |
|----|--|---------------------------|
| | RESULT | 8 |
| | AAM21759 | |
| ID | AAM21759 | standard; protein; 63 AA. |
| XX | | |
| AC | AAM21759; | |
| DT | 12-OCT-2001 | (first entry) |
| XX | | |
| DE | Peptide #8193 encoded by probe for measuring cervical gene expression. | |
| XX | Probe; human; microarray; gene expression; cervical epithelial cell; | |

| | | |
|----|----------------------------------|--------------------------------|
| | RESULT | 7 |
| XX | AAR13320 | |
| ID | AAR13320 | standard; Protein; 982 AA. |
| XX | | |
| AC | AAR13320; | |
| XX | | |
| DT | 22-OCT-1991 | (first entry) |
| XX | | |
| DE | Murine Natural Killer receptor. | |
| XX | | |
| KM | NK cytotoxic drugs; tumour cell. | |
| XX | | |
| OS | Homo sapiens. | |
| XX | | |
| FH | Key | Location/Qualifiers |
| FT | Protein | 8..982 |
| FT | | /label= mature_NK_receptor |
| FT | Peptide. | 1..7 |
| XX | | /label= partial_signal_peptide |
| PN | US7535206-A. | |
| XX | | |
| PD | 09-JUL-1991. | |
| XX | | |

| | |
|----|---|
| OS | Homo sapiens. |
| XX | |
| PN | MO200157278-A2. |
| XX | |
| PD | 09-AUG-2001. |
| XX | |
| PE | 30-JAN-2001; 2001WO-US00670. |
| XX | |
| PR | 04-FEB-2000; 2000US-0180312. |
| PR | 26-MAY-2000; 2000US-0207456. |
| PR | 30-JUN-2000; 2000US-0608408. |
| PR | 03-AUG-2000; 2000US-0632366. |
| PR | 21-SEP-2000; 2000US-0234687. |
| PR | 27-SEP-2000; 2000US-0236359. |
| PR | 04-OCT-2000; 2000GB-0024253. |
| XX | |
| PA | {MOLE-} MOLECULAR DYNAMICS INC. |
| XX | |
| PI | Penn SG, Hanzel DK, Chen W, Rank DR; |
| XX | |
| DR | WPI; 2001-488901/53. |
| XX | |
| PT | Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human cervical epithelial cells - |

| | | |
|-----------|---|---|
| PS | XX | Claim 27; SEQ ID No 26585; 487bp; English. |
| CC | XX | |
| CC | CC | The present invention relates to human single exon nucleic acid probes (SNP: see AA110068-AA128459). The present sequence is a peptide encoded by one such probe. The SEMP are derived from human Hela cells. The SEMP can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer. |
| CC | CC | Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences. |
| CC | CC | |
| CC | XX | |
| SO | Sequence | 63 AA: |
| QY | Query Match | 28.8%; Score 72.5; DB 22; Length 63; |
| Db | Best Local Similarity | 43.1%; Pred. No. 0.091; Mismatches 20; Indels 5; Gaps 2; |
| | Matches 22; Conservative | 4; Mismatches 20; Indels 5; Gaps 2; |
| | 1 | STGSHGSGSHHTTGGKSDASRCG-SGSRSTSRRETRDQSGDGSG---RH 46 |
| | 5 | stssgssststgssgsssssaarsgrsgssstarrsssssssgpsprtrrh 55 |
| RESULT | 9 | |
| AAAM38081 | ID | AAAM38081 standard; Protein; 63 AA. |
| XX | XX | |
| XX | AAAM38081; | |
| DT | 17-OCT-2001 | (first entry) |
| XX | XX | |
| DE | Peptide #12118 encoded by probe for measuring placental gene expression. | |
| XX | XX | |
| KW | Probe; microarray; human; placenta; antenatal diagnosis; | |
| KW | genetic disorder. | |
| XX | XX | |
| XX | Homo sapiens. | |
| PN | WO200157272-A2. | |
| PD | XX | |
| PD | 09-AUG-2001. | |
| PF | 30-JAN-2001; 2001WO-US00663. | |
| XX | XX | |
| XX | 04-FEB-2000; 2000US-0180312. | |
| PR | 26-MAY-2000; 2000US-0207456. | |
| PR | 30-JUN-2000; 2000US-0608408. | |
| PR | 03-AUG-2000; 2000US-0632366. | |
| PR | 21-SEP-2000; 2000US-0234687. | |
| PR | 27-SEP-2000; 2000US-0236359. | |
| PR | 04-OCT-2000; 2000GB-0024263. | |
| XX | XX | |
| PA | (MOLE-) MOLECULAR DYNAMICS INC. | |
| XX | XX | |
| PI | Penn SG, Hanzel DK, Chen W, Rank DR; | |
| DR | WPI: 2001-488897/53. | |
| XX | XX | |
| PT | Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human placenta - | |
| XX | XX | |
| PS | Claim 27; SEQ ID No 38350; 654bp; English. | |
| CC | XX | |
| CC | CC | The present invention relates to single exon nucleic acid probes (SNP: see AA131315-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders. |

| | | |
|----|---|--|
| XX | Sequence | 63 AA; |
| SQ | | |
| | Query Match | 28.8%; Score 72.5; DB 22; Length 63; |
| | Best Local Similarity | 43.1%; Pred. No. 0.091; |
| | Matches 22; Conservative | 4; Mismatches 20; Indels 5; Gaps 2; |
| OY | 1 | STGHSGSCHSHTTTGGKSDASRGSGSRSTRSRETRDOEGSDGS----RH 46 |
| | : : : | |
| Dd | 5 | stssgssstctsgsssgssssasssrsgsstcrsssssssgsparrhh 55 |
| | | |
| | RESULT 10 | |
| | AAB93496 | |
| ID | AAB93496 standard; Protein: | 407 AA. |
| XX | | |
| AC | AAB93496; | |
| XX | | |
| DT | 26-JUN-2001 | (first entry) |
| XX | | |
| DE | Human protein sequence SEQ ID NO:12809. | |
| KW | Human; primer: detection; diagnosis; antisense therapy; gene therapy. | |
| XX | Homo sapiens. | |
| OS | | |
| XX | Epi074617-A2. | |
| FN | | |
| PD | 07-FEB-2001. | |
| XX | | |
| PF | 28-JUL-2000; 2000EP-0116126. | |
| XX | | |
| PR | 29-JUL-1999; 99JP-0248036. | |
| PR | 27-AUG-1999; 99JP-0300253. | |
| PR | 11-JAN-2000; 2000JP-0118776. | |
| PR | 02-MAY-2000; 2000JP-0183767. | |
| PR | 09-JUN-2000; 2000JP-0241899. | |
| XX | | |
| PA | (HELI-) HELIX RES INST. | |
| XX | | |
| PI | Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J; | |
| PL | Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T; | |
| XX | | |
| DR | WPI; 2001-318749/34. | |
| XX | | |
| Pt | Primer sets for synthesizing polynucleotides, particularly the 5602 | |
| Pt | full-length cDNAs defined in the specification, and for the detection | |
| Pt | and/or diagnosis of the abnormality of the proteins encoded by the | |
| Pt | full-length cDNAs - | |
| PS | | |
| PS | Claim 8; SEQ ID 12809; 2537pp + CD ROM; English. | |
| XX | | |
| CC | The present invention describes primer sets for synthesizing 5602 | |
| CC | full-length cDNAs defined in the specification. Where a primer set | |
| CC | comprises: (a) an oligo-dT primer and an oligonucleotide complementary | |
| CC | to the complementary strand of a polynucleotide which comprises one of | |
| CC | the 5602 nucleotide sequences defined in the specification, where the | |
| CC | oligonucleotide comprises at least 15 nucleotides; or (b) a combination | |
| CC | of an oligonucleotide comprising a sequence complementary to the | |
| CC | complementary strand of a polynucleotide which comprises a 5'-end | |
| CC | sequence and an oligonucleotide comprising a sequence complementary to a | |
| CC | polynucleotide which comprises a 3'-end sequence, where the | |
| CC | oligonucleotide comprises at least 15 nucleotides and the combination of | |
| CC | the 5'-end sequence/3'-end sequence is selected from those defined in | |
| CC | the specification. The primer sets can be used in antisense therapy and | |
| CC | in gene therapy. The primers are useful for synthesizing polynucleotides, | |
| CC | particularly full-length cDNAs. The primers are also useful for the | |
| CC | detection and/or diagnosis of the abnormality of the proteins encoded by | |
| CC | the full-length cDNAs. The primers allow obtaining of the full-length | |
| CC | cDNAs easily without any specialised methods. AAH03166 to AAH13628 and | |
| CC | AAH13693 to AAH18742 represent human cDNA sequences; AAB92446 to | |
| CC | AAB95983 represent human amino acid sequences; and AAH13629 to AAH13632 | |
| CC | | |

| | |
|------|---|
| Pt | Kaufmann J; |
| Xx | |
| Df | WPI; 2000-116524/10. |
| Nr | N-PSSB; AAZ29604. |
| Xx | |
| Pt | Novel regulatory protein, useful to diagnose and treat cell proliferation disorders, including cancer - |
| Ft | |
| Xx | Claim 1; Fig 1; 52pp; English. |
| Xx | |
| Cc | The present sequence is one form of human Cifl30 protein which binds to Ciffl50/HtaFlil50 and inhibits mitosis or cell cycle progression. Cifl30 shows homology to ATP-dependent RNA helicase (Dead-box protein) and is differentially expressed in tumours, such as human brain, uterine, Fallopian tube and ovarian tumours, compared to normal tissues. The present sequence is used to diagnose and treat proliferative disorders including hyperplasiae, neoplasias and dysplasias. The sequence can also be used to screen compounds that interfere with binding of Cifl30 to Cifl50/HtaFlil50 protein. |
| Cc | |
| Sq | Sequence 1275 AA; |
| Oy | Query Match 28.0%; Score 70.5; DB 21; Length 1275; Best Local Similarity 34.1%; Pred. No. 4.5; Matches 15; Conservative 8; Mismatches 16; Indels 5; Gaps 1. 3 GHSGSHSTTTQGRSDASRGSSGSRSTRERRDQGSGSRRH 46 : : : : : : : : : Db gsngrkrerytenrgsrnshgctgnrh-----dsprngdgqrh 885 |
| Rslt | _13 |
| ID | AABJ3581 |
| Xx | AABJ3581 standard; Protein; 717 AA. |
| Ac | |
| Xx | AABJ3581; |
| Df | |
| Xx | 20-APR-2001 (first entry) |
| De | |
| Xx | Amino acid sequence of a human plnin splice variant (PNIN). |
| Xx | |
| Kw | Human; plnin splice variant; PNIN; vesicle trafficking; AIDS; emphysema; Addison's disease; neoplastic disorder; immunological disorder; gout; adult respiratory distress syndrome; allergy; anaemia; asthma; cancer; atherosclerosis; bronchitis; cholecyellitis; Crohn's disease; scleroderma; ulcerative colitis; atopic dermatitis; dermatomyositis; osteoporosis; atrophic gastritis; diabetes mellitus; erythema nodosum; pancreatitis; glomerulonephritis; Graves' disease; hypereosinophilia; polymyositis; irritable bowel syndrome; lupus erythematosus; multiple sclerosis; myasthenia gravis; myocardial or pericardial inflammation; hemodialysis; osteoarthritis; rheumatoid arthritis; Sjogren's syndrome; trauma; Werner syndrome; autoimmune thyroiditis; infection. |
| Kv | |
| Rw | Homo sapiens. |
| Xx | |
| Os | |
| Xx | |
| Fk | Key |
| Ft | Modified-site Location/Qualifiers |
| Ft | /note= "potential casein kinase II phosphophorylation site" |
| Ft | Modified-site 18 |
| Ft | /note= "potential protein kinase C phosphorylation site" |
| Ft | Modified-site 30 |
| Ft | /note= "potential protein kinase C phosphorylation site" |
| Ft | Modified-site 30 |
| Ft | /note= "potential casein kinase II phosphophorylation site" |
| Ft | Modified-site 30 |
| Ft | /note= "potential cAMP- and cGMP-dependent protein kinase phosphorylation site" |
| Ft | Modified-site 92 |
| Ft | /note= "potential casein kinase II phosphophorylation site" |

| | | | |
|----|---------------|-----|--|
| FT | Modified-site | 92 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 96 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 96 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 96 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 96 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 100 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 100 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 119 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 119 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 119 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 124 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 124 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 124 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 164 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 252 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 259 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 289 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 289 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 289 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 289 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 289 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 375 | "potential CAMP- and cGMP-dependent protein kinase phosphorylation site" |
| FT | Modified-site | 408 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 417 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 417 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 448 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 450 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 500 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 500 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 534 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 568 | "potential glycosylation site" |
| FT | /note= | 570 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 602 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 617 | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 617 | "potential protein kinase C phosphorylation site" |
| FT | /note= | 617 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 627 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 654 | "potential CAMP- and cGMP-dependent protein kinase phosphorylation site" |
| FT | Modified-site | 658 | "potential casein kinase II phosphos phosphorylation site" |
| FT | /note= | 671 | "potential casein kinase II phosphos phosphorylation site" |
| FT | Modified-site | 671 | "potential casein kinase II phosphos phosphorylation site" |

| | | | |
|----|---|-------------------|--|
| FT | Modified-site | /note= | "potential casein kinase II phosphosphorylation site" |
| FT | Modified-site | 663 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 666 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 671 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 683 | |
| FT | Modified-site | /note= | "potential casein kinase II phosphosphorylation site" |
| FT | Modified-site | 685 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 687 | |
| FT | Modified-site | /note= | "potential cAMP- and cGMP-dependent protein kinase phosphorylation site" |
| FT | Modified-site | 690 | |
| FT | Modified-site | /note= | "potential casein kinase II phosphosphorylation site" |
| FT | Modified-site | 694 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 697 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 699 | |
| FT | Modified-site | /note= | "potential cAMP- and cGMP-dependent protein kinase phosphorylation site" |
| FT | Modified-site | 701 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 702 | |
| FT | Modified-site | /note= | "potential casein kinase II phosphosphorylation site" |
| FT | Modified-site | 704 | |
| FT | Modified-site | /note= | "potential casein kinase II phosphosphorylation site" |
| FT | Modified-site | 704 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| FT | Modified-site | 710 | |
| FT | Modified-site | /note= | "potential protein kinase C phosphorylation site" |
| XX | US6162601-A. | | |
| XX | 19-DEC-2000. | | |
| PD | 08-AUG-1997; | 97US-0910925. | |
| XX | 08-AUG-1997; | 97US-0910925. | |
| XX | 08-AUG-1997; | 97US-0910925. | |
| XX | (INCY-) | INCYTE PHARM INC. | |
| XX | Bandman O, Shah P, Lal P; | | |
| XX | WPI: 2001-090399/10. | | |
| DR | N-PSDB; AAF24910. | | |
| XX | Novel nucleotide sequence encoding human pinin splice variant, useful for treating AIDS, Addison's disease, allergy, anaemia, asthma, atherosclerosis and adult respiratory distress syndrome | | |
| XX | Claim 1; Fig 1A-G; 38pp; English. | | |
| PS | The present sequence represents a human pinin splice variant (PININ). PININ | | |
| CC | polypeptides and polynucleotides are useful for treating, diagnosing and | | |
| CC | preventing developmental, vesicle trafficking, neoplastic and | | |
| CC | immunological disorders. Antagonists of PININ polypeptides are useful | | |
| CC | for treating AIDS. Addison's disease, adult respiratory distress | | |
| CC | syndrome, allergies, anaemia, asthma, atherosclerosis, bronchitis, | | |
| CC | cholecystitis, Crohn's disease, ulcerative colitis, atopic dermatitis, | | |
| CC | chemodectomas, diabetes mellitus, emphysema, erythema nodosum, atrophic | | |
| CC | gastritis, glomerulonephritis, gout, Graves' disease, hyperosinophilia, | | |
| CC | irritable bowel syndrome, lupus erythematosus, multiple sclerosis, | | |
| CC | myasthenia gravis, myocardial or pericardial inflammation, | | |
| CC | osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid | | |
| CC | arthritis, scleroderma, Sjogren's syndrome, Werner syndrome, and | | |

| | |
|-----------|---|
| CC | autoimmune thyroiditis, complications of cancer, hemodialysis, and |
| CC | extracorporeal circulation, trauma, and viral, bacterial fungal, |
| CC | parasitic, protozoal, and helminthic infections. |
| XX | |
| SO | Sequence 717 AA; |
| QY | 1 STGHSQSHHTTTCGRSDASGSGSRSTSTETRDQESGSGSRHSGS 49 |
| Db | 577 sssssssssstsssssgsssssgsssssrss--ssssstgssssrdsss 623 |
| RESULT 14 | |
| AAB54149 | |
| ID | AAB54149 standard; Protein: 306 AA. |
| XX | |
| AC | AAB54149; |
| XX | |
| DT | 09-MAR-2001 (first entry) |
| XX | |
| DE | Human pancreatic cancer antigen protein sequence SEQ ID NO:601. |
| XX | |
| KW | Human; pancreas; pancreatic cancer; pancreatic cancer antigen; |
| KW | detection; diagnosis; identification; cytostatic; neuroprotective; |
| KW | neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological; |
| KW | antiinflammatory; cardiant; gene therapy; chromosome mapping; |
| KW | linkage analysis; tissue identification; tissue typing; forensic; |
| KW | neural; immune system; muscular; reproductive; gastrointestinal; |
| XX | pulmonary; cardiovascular; renal; proliferative. |
| XX | |
| OS | Homo sapiens. |
| XX | |
| PN | WO200055320-A1. |
| XX | |
| PD | 21-SEP-2000. |
| XX | |
| PF | 08-MAR-2000; 2000WO-US05989. |
| XX | |
| PR | 12-MAR-1999; 99US-0124270. |
| XX | |
| PA | (HUMA-) HUMAN GENOME SCI INC. |
| XX | |
| PI | Rosen CA, Ruben SM; |
| XX | |
| DR | WPI: 2000-579444/54. |
| XX | |
| DR | N-PSDB; AAC98914. |
| XX | |
| PT | New nucleic acid that is a pancreatic cancer antigen for preventing, |
| PT | treating, or ameliorating a medical condition, particular pancreatic |
| PT | cancer, or for use in assays for diagnosing a pathological condition - |
| XX | |
| XX | Claim 11; Page 1037-1038; 1379pp; English. |
| PS | |
| CC | AAC98773 to AAC99231 encode the human pancreatic cancer associated |
| CC | proteins, called pancreatic cancer antigens, given in AAB5408 to |
| CC | AAB54466. The human pancreatic cancer antigens have cytostatic, |
| CC | neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive, |
| CC | gynaecological, cardiant and antiinflammatory activities, and can be used |
| CC | in gene therapy. The polynucleotide and proteins can be used for |
| CC | preventing, treating, or ameliorating a medical condition or in assays |
| CC | for diagnosing a pathological condition or a susceptibility to one in a |
| CC | subject. Binding partners to the proteins and the activity of the |
| CC | proteins can be identified. The pancreatic cancer antigens can be used to |
| CC | detect, treat or prevent pancreatic disorders, especially cancer. |
| CC | Agonists and antagonists to the antigens can be screened for. The |
| CC | pancreatic cancer antigen polynucleotides can be used to design nucleic |
| CC | acid hybridisation probes that can be used in chromosome mapping, linkage |
| CC | analysis, tissue identification and/or typing and a variety of forensic |
| CC | and diagnostic methods. The proteins can be used to generate antibodies |
| CC | |

CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.

XX Sequence 306 AA;

Query Match 26.6%; Score 67; DB 21; Length 306;
Best Local Similarity 37.0%; Pred. No. 2.4;
Matches 20; Conservative 9; Mismatches 17; Indels 8; Gaps 2;

OY 2 TGHSGSQSHSTTTTCGRSDA-----SRGSSGSRSTSRRTPDQSGDGR--HS 47
Db 140 sgsskshsksrstrs 193

RESULT 15

AAM40069
ID AAM40069 standard; Protein: 443 AA.

AC AAM40069;

DT 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 3214.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX MPI: 2001-442253/47.

DR N-PSDB; AAI59225.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX PT such as central nervous system injuries -

XX PS Example 5; SEQ ID NO 3214; 10078bp; English.

CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 443 AA;

Query Match 26.6%; Score 67; DB 22; Length 443;
Best Local Similarity 38.9%; Pred. No. 3.6;
Matches 21; Conservative 8; Mismatches 19; Indels 6; Gaps 3;

OY 1 STGHSGSQSHSTTT--QGRSDASRSGSRSTSRRTPDQSGC--DGRHSGS 49
Db 377 ntssnrtrsgswsggrsggrsggrsggrs-grgsrgrsgrsgrdgrrrsgn 429

Search completed: January 28, 2002, 07:39:28
Job time: 118 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 ; Search time 12.62 Seconds
(without alignments)
87.374 Million cell updates/sec

Title: US-09-582-711-3

Perfect score: 252
Sequence: 1 STGSGSHSHHTTQGRSDA.....TSRETRDQSGSGSRHSGS 49

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------------|
| 1 | 73 | 29.0 | 743 | 4 | US-08-910-925-3 |
| 2 | 70.5 | 28.0 | 1261 | 4 | US-09-208-742-4 |
| 3 | 70.5 | 28.0 | 1261 | 4 | US-09-332-295-2 |
| 4 | 69.5 | 27.6 | 703 | 4 | US-08-910-925-4 |
| 5 | 69 | 27.4 | 717 | 4 | US-08-910-925-1 |
| 6 | 65 | 25.8 | 586 | 3 | US-08-964-268-3 |
| 7 | 64 | 25.4 | 3025 | 6 | 5223423-3 |
| 8 | 61.5 | 24.4 | 1093 | 3 | US-08-545-860D-55 |
| 9 | 61.5 | 24.4 | 1093 | 5 | PCT-US94-04496-55 |
| 10 | 61 | 24.2 | 1581 | 4 | US-09-110-517-2 |
| 11 | 60 | 23.8 | 984 | 1 | US-08-257-073-3 |
| 12 | 60 | 23.8 | 984 | 2 | US-08-184-009-120 |
| 13 | 60 | 23.8 | 984 | 2 | US-08-458-356-120 |
| 14 | 60 | 23.8 | 984 | 4 | US-08-460-736-120 |
| 15 | 57.5 | 22.8 | 514 | 4 | US-08-796-899-25 |
| 16 | 57 | 22.6 | 414 | 1 | US-07-667-276A-4 |
| 17 | 56.5 | 22.4 | 73 | 1 | US-08-272-882D-3 |
| 18 | 56.5 | 22.4 | 73 | 5 | PCT-US92-05532-3 |
| 19 | 56.5 | 22.4 | 1481 | 2 | US-08-616-844-40 |
| 20 | 56.5 | 22.4 | 1481 | 2 | US-08-599-654-40 |
| 21 | 56.5 | 22.4 | 1481 | 3 | US-08-944-868A-40 |
| 22 | 56.5 | 22.4 | 1481 | 3 | US-08-944-423A-40 |
| 23 | 56.5 | 22.4 | 1481 | 3 | US-08-944-496-40 |
| 24 | 56.5 | 22.4 | 3969 | 4 | US-08-061-376-5 |
| 25 | 56 | 22.2 | 376 | 4 | US-09-125-099-10 |
| 26 | 56 | 22.2 | 376 | 4 | US-09-125-099-12 |
| 27 | 56 | 22.2 | 400 | 5 | PCT-US95-16472-2 |

| | | | | | | |
|----|------|------|-----|---|-------------------|--------------------|
| 28 | 56 | 22.2 | 402 | 3 | US-08-602-809-2 | Sequence 2, Appl1 |
| 29 | 56 | 22.2 | 568 | 1 | US-08-320-559-30 | Sequence 30, Appl1 |
| 30 | 56 | 22.2 | 568 | 3 | US-08-545-860D-30 | Sequence 30, Appl1 |
| 31 | 56 | 22.2 | 568 | 5 | PCT-US94-04496-30 | Sequence 30, Appl1 |
| 32 | 55.5 | 22.0 | 401 | 2 | US-08-549-004A-5 | Sequence 5, Appl1 |
| 33 | 55.5 | 22.0 | 401 | 4 | US-09-051-982A-5 | Sequence 10, Appl1 |
| 34 | 55.5 | 22.0 | 410 | 2 | US-08-723-415B-10 | Sequence 11, Appl1 |
| 35 | 55.5 | 22.0 | 410 | 2 | US-08-723-415B-11 | Sequence 2, Appl1 |
| 36 | 55.5 | 22.0 | 410 | 2 | US-08-428-131-2 | Sequence 2, Appl1 |
| 37 | 55.5 | 22.0 | 410 | 2 | US-08-602-846-2 | Sequence 2, Appl1 |
| 38 | 55.5 | 22.0 | 410 | 4 | US-09-078-596-2 | Sequence 2, Appl1 |
| 39 | 55.5 | 22.0 | 410 | 4 | US-09-189-627A-10 | Sequence 10, Appl1 |
| 40 | 55.5 | 22.0 | 410 | 4 | US-09-189-627A-11 | Sequence 11, Appl1 |
| 41 | 55.5 | 22.0 | 763 | 2 | US-08-677-862-2 | Sequence 2, Appl1 |
| 42 | 55.5 | 22.0 | 763 | 2 | US-09-252-571-2 | Sequence 2, Appl1 |
| 43 | 55.5 | 22.0 | 763 | 3 | US-09-434-065-2 | Sequence 2, Appl1 |
| 44 | 55.5 | 22.0 | 763 | 4 | US-08-789-275-4 | Sequence 4, Appl1 |
| 45 | 55.5 | 22.0 | 763 | 4 | US-08-789-275-5 | Sequence 5, Appl1 |

ALIGNMENTS

RESULT 1
US-08-910-925-3
; Sequence 3, Application US/08910925
; Patent No. 6162601
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OR INVENTION: HUMAN PININ SPLICE VARIANT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,925
; FILING DATE: Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0365 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 743 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1684847
; US-08-910-925-3
Query Match 29.0%; Score 73; DB 4; Length 743;

[illegible]

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-964-268-3

Query Match      25.8%; Score 65; DB 3; Length 586;
Best Local Similarity 41.2%; Pred. No. 2.7;
Matches 14; Conservative 3; Mismatches 17; Indels 0; Gaps 0;

OY   4 HSGSHSTTTQGRDASRGSSGSRSTRERDQ 37
      |  || :| |||| |  || :| :|
Db    541 HGDSQOTHLLKQGRSSMGTGLSGKRPSQEEDTQ 574

RESULT       7
5223423-3
; Patent No. 5223423
; APPLICANT: FRANCHINI, GENOVEPPA,WONG-STRAAL, FLOSSIE;
; GALLO, ROBERT
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/331,212
; FILING DATE: 03-31-1989
; SEQ ID NO.:3:
; LENGTH: 3025
5223423-3

Query Match      25.4%; Score 64; DB 6; Length 3025;
Best Local Similarity 27.0%; Pred. No. 24;
Matches 17; Conservative 7; Mismatches 21; Indels 18; Gaps 1;

OY   5 SCGSHSTTTQGRSDASRGSSGSRSTRERTR-----DOEGSGDGSRH 46
      :| || | :| :| ||| :| :| :| :| :| :| :| :| :| :| :|
Db    2644 TSSNSETGGDSNQPCSMGASGSKRKRSPRGDLRLLRANGACGGLMDSEGGYSQFH 2703

OY   47 SGS 49
      ||
Db    2704 EGS 2706

RESULT       8
US-08-545-860D-55
; Sequence 55, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; APPLICANT: CanaanI, Eli
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
; TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
; ADDRESSEE: No. 6040140rIs
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: USA
ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,860D

```

```

1      FILING DATE: 07-JAN-1996
2      CLASSIFICATION: 435
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER: PCT/US94/04496
5      FILING DATE: 22-APR-1994
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER: PCT/US92/10930
8      FILING DATE: 09-DEC-1992
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER: US 08/327,392
11     FILING DATE: 19-OCT-1994
12     PRIOR APPLICATION DATA:
13     APPLICATION NUMBER: US 08/320,559
14     FILING DATE: 11-OCT-1994
15     PRIOR APPLICATION DATA:
16     APPLICATION NUMBER: US 08/062,443
17     FILING DATE: 14-MAY-1993
18     PRIOR APPLICATION DATA:
19     APPLICATION NUMBER: US 07/971,094
20     FILING DATE: 30-OCT-1992
21     PRIOR APPLICATION DATA:
22     APPLICATION NUMBER: US 07/888,839
23     FILING DATE: 27-MAY-1992
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER: US 07/805,093
26     FILING DATE: 11-DEC-1991
27     ATTORNEY/AGENT INFORMATION:
28     NAME: Deluca Esq., Mark
29     REGISTRATION NUMBER: 33,229
30     REFERENCE/DOCKET NUMBER: TU-1262
31     TELECOMMUNICATION INFORMATION:
32     TELEPHONE: (215) 568-3100
33     TELEFAX: (215) 568-3439
34     INFORMATION FOR SEQ ID NO: 55:
35     SEQUENCE CHARACTERISTICS:
36     LENGTH: 1093 amino acids
37     TYPE: amino acid
38     STRANDEDNESS: single
39     TOPOLOGY: linear
40     MOLECULE TYPE: Protein
41     HYDROTHERICAL: NO
42     ANTI-SENSE: NO
43     US-08-545-860D-55

```

```

      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.25
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: PCT/US94/04496
      FILING DATE:
      CLASSIFICATION:
      ATTORNEY/AGENT INFORMATION:
      NAME: Deluca Esq., Mark
      REGISTRATION NUMBER: 33,229
      REFERENCE/DOCKET NUMBER: TJU-1242
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (215) 568-3100
      TELEFAX: (215) 568-3439
      INFORMATION FOR SEQ ID NO: 55:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 1093 amino acids
      TYPE: amino acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: protein
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
PCT-US94-04496-55

```

```

Query Match      24.4%; Score 61.5; DB 3; Length 1093;
Best Local Similarity 34.8%; Pred. No. 15;
Matches 16; Conservative 8; Mismatches 19; Indels 3; Gaps 1;

      1  STGHSQSQSHSTTTQGRSDASRGSSGSRSTSRHTRDOEO---SGDG 43
      |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      275  SSSASSSSHHASSTQETSESSRESNGKSSSSLSHKGKRLSSGKG 320

RESULT          9
PCT-US94-04496-55
: Sequence 55, Application PC/TUS9404496
:
: GENERAL INFORMATION:
:
: APPLICANT: Croce, Carlo
: APPLICANT: Canaan, Eli
: TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
: TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
: TITLE OF INVENTION: Resulting from Chromosome Abnormalities In the A11-1
: NUMBER OF SEQUENCES: 86
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Woodcock, Washburn, Kurtz, Macklewitz &
: ADDRESSEE: Norris
: STREET: One Liberty Place, 46th floor
: CITY: Philadelphia
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19103
:
: COMPUTER READABLE FORM:

```

```

OY      1  STHGSGQSHHTTQG-RSDASGRSSGSRSTRETRODQEQ--S GGG 43
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      275 SSSASSSSHHEASTQETSESSRESKXKKSSHSLSKCKKLSSGKG 320
Matches 16; Conservative 8; Mismatches 19; Indels 3; Gaps 1

RESULT 10
US-09-110-517-2
; Sequence 2, Application US/09110517A
; Patent No. 6248520
; GENERAL INFORMATION:
; APPLICANT: Roeder, Robert G
; APPLICANT: Fondell, Joseph D
; APPLICANT: Yuan, Chao X
; APPLICANT: Ito, Mitsuhiko
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NUCLEAR HORMONE
; TITLE OF INVENTION: RECEPTOR COACTIVATORS AND USES THEREOF
; FILE REFERENCE: 600-1-224
; CURRENT APPLICATION NUMBER: US/09/110.517A
; CURRENT FILING DATE: 1998-07-06
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1581
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-110-517-2

Query Match          24.2%; Score 61; DB 4; Length 1581;
Best Local Similarity 35.4%; Pred. No. 26;
Matches 17; Conservative 5; Mismatches 24; Indels 2; Gaps 1;

OY      4  HSGSQSHHTTQGG--RSDASRGSSGSRSTRETRODQEGDGSGSRHSGS 49
       ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     1100 HSSSSSSASTSGMKMSKSECGSSSKLSSMYSSQGSSSQSQKNSS 1147

RESULT 11
US-08-257-073-3
; Sequence 3, Application US/08257073
; Patent No. 576597
GENERAL INFORMATION:
```

```

1 APPLICANT: Paoletti, Enzo
2 APPLICANT: de Taisne, Charles
3 APPLICANT: Tine, John A.
4 TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
5 NUMBER OF SEQUENCES: 143
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: Curtiss, Morris & Safford, P.C.
8 STREET: 530 Fifth Avenue, 25th Floor
9 CITY: New York
10 STATE: New York
11 COUNTRY: UNITED STATES OF AMERICA
12 ZIP: 10036
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentln Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/257,073
20 FILING DATE: 09-JUN-1994
21 CLASSIFICATION: 424
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 08/075,783
24 FILING DATE: 11-JUN-1993
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 07/852,305
27 FILING DATE: 18-MAR-1992
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 07/672,183
30 FILING DATE: 20-MAR-1991
31 ATTORNEY/AGENT INFORMATION:
32 NAME: Frommer, William S.
33 REGISTRATION NUMBER: 25,506
34 REFERENCE/DOCKET NUMBER: 454310-2570
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: (212) 840-3333
37 TELEFAX: (212) 840-0712
38 TEXAS: 425066 CURTMS
39 INFORMATION FOR SEQ ID NO: 3:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 984 amino acids
42 TYPE: amino acid
43 STRANDEDNESS: single
44 TOPOLOGY: linear
45 MOLECULE TYPE: peptide
46 FRAGMENT TYPE: internal
47 US-08-257-073-3

```

```

1 COUNTRY: USA
2 ZIP: 10036
3
4 COMPUTER READABLE FORM:
5 MEDIUM TYPE: Floppy disk
6 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOS
8 SOFTWARE: Patentin Release #1.0, Version #1.25
9 CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/08/184,009
11 FILING DATE: 19-JAN-1994
12 CLASSIFICATION: 435
13 ATTORNEY/AGENT INFORMATION:
14 NAME: Frommer, William S.
15 REGISTRATION NUMBER: 25,506
16 REFERENCE/DOCKET NUMBER: 454310-2530
17 TELECOMMUNICATION INFORMATION:
18 TELEPHONE: (212) 840-3333
19 TELEFAX: (212) 840-0712
20 TELEX: 425066CURPMS
21 INFORMATION FOR SEQ ID NO: 120:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 984 amino acids
24 TYPE: amino acid
25 STRANDEDNESS: single
26 TOPOLOGY: linear
27 MOLECULE TYPE: peptide
28 FRAGMENT TYPE: internal
29
30 US-08-184-009-120
31
32 Query Match 23.8% Score 60; DB 2; Length 984;
33 Best Local Similarity 31.8% Pred. NO. 20;
34 Matches 14; Conservative 8; Mismatches 22; Indels 0; Gaps 0;
35
36 Oy 1 STGHSGSOHSHTTGTGRSDASRGSGSRSTSRRTDQSGGGS 44
37 Db 175 SNGTTGEOSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 218
38
39 RESULT 13
40 US-08-458-356-120
41 Sequence 120 Application US/08458356
42 Patent No. 5942235
43 GENERAL INFORMATION:
44 APPLICANT: Paolelli, Enzo
45 APPLICANT: Paolelli, James
46 APPLICANT: Cox, William I.
47 TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
48 NUMBER OF SEQUENCES: 217
49 CORRESPONDENCE ADDRESS:
50 ADDRESSEE: Curtiss, Morris & Safford
51 STREET: 530 Fifth Avenue
52 CITY: New York
53 STATE: NY
54 COUNTRY: USA
55 ZIP: 10036
56
57 COMPUTER READABLE FORM:
58 MEDIUM TYPE: Floppy disk
59 COMPUTER: IBM PC compatible
60 OPERATING SYSTEM: PC-DOS/MS-DOS
61 SOFTWARE: Patentin Release #1.0, Version #1.25
62 CURRENT APPLICATION DATA:
63 APPLICATION NUMBER: US/08/458,356
64 FILING DATE: 02-JUN-1995
65 CLASSIFICATION: 424
66 PRIOR APPLICATION DATA:
67 APPLICATION NUMBER: US 08/184,009
68 FILING DATE: 19-JAN-1994
69 ATTORNEY/AGENT INFORMATION:
70 NAME: Frommer, William S.
71 REGISTRATION NUMBER: 25,506
72 REFERENCE/DOCKET NUMBER: 454310-2530
73 TELECOMMUNICATION INFORMATION:

```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 : Search time 13.52 Seconds
(without alignments)
276.076 Million cell updates/sec

Title: US-09-582-711-3

Perfect score: 252
Sequence: 1 STGSGSQSHHTTQGRSDA.....TSRETRDQSGSGSRHSGS 49

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*\n2: pir2:*\n3: pir3:*\n4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 252 | 100.0 | 2248 | 2 A35938 | profilaggrin - hum |
| 2 | 207 | 82.1 | 416 | 2 A32947 | filaggrin precursor |
| 3 | 207 | 82.1 | 591 | 2 A45135 | profilaggrin - hum |
| 4 | 77 | 30.6 | 343 | 2 T05221 | hypothetical prote |
| 5 | 73.5 | 29.2 | 716 | 2 S43693 | penicillin-binding |
| 6 | 73.5 | 29.2 | 1507 | 2 B47328 | natural killer cel |
| 7 | 72.5 | 28.8 | 305 | 2 TC4525 | nucleic acid-bindi |
| 8 | 71 | 28.2 | 196 | 2 T29343 | hypothetical prote |
| 9 | 70.5 | 28.0 | 819 | 2 T08745 | probable RNA helic |
| 10 | 69.5 | 27.6 | 471 | 2 T33997 | hypothetical prote |
| 11 | 69.5 | 27.6 | 1217 | 2 S52714 | sericidinb - silkw |
| 12 | 69 | 27.4 | 174 | 2 S18654 | variant surface an |
| 13 | 69 | 27.4 | 284 | 2 A35419 | neutrophil protein |
| 14 | 69 | 27.4 | 625 | 2 A34615 | profilaggrin - rat |
| 15 | 68.5 | 27.2 | 254 | 2 A31488 | filaggrin - mouse |
| 16 | 68.5 | 27.2 | 313 | 2 A28444 | filaggrin precursor |
| 17 | 68 | 27.0 | 406 | 2 S38170 | SRP40 protein - ye |
| 18 | 68 | 27.0 | 1217 | 2 T26255 | AF-4 protein - mou |
| 19 | 67.5 | 26.8 | 156 | 2 H34768 | ORF8 protein - Orf |
| 20 | 67 | 26.6 | 399 | 2 T46239 | hypothetical prote |
| 21 | 67 | 26.6 | 494 | 2 A48133 | pre-mRNA splicing |
| 22 | 67 | 26.6 | 669 | 2 T44681 | GTP-binding protei |
| 23 | 66.5 | 26.4 | 635 | 2 F75477 | hypothetical prote |
| 24 | 66 | 26.2 | 860 | 2 JC5702 | ErkB kinase activa |
| 25 | 66 | 26.2 | 868 | 2 JC5701 | ErkB kinase activa |
| 26 | 65.5 | 26.0 | 115 | 2 D61615 | sericin Mg-1 - gre |
| 27 | 65.5 | 26.0 | 294 | 2 T35892 | hypothetical prote |
| 28 | 65.5 | 26.0 | 1207 | 2 T23754 | hypothetical prote |
| 29 | 65.5 | 26.0 | 1469 | 2 T09219 | basal transcriptio |

| | | | | | |
|----|------|------|------|----------|--------------------|
| 30 | 65 | 25.8 | 635 | 2 T00011 | ccal protein - rat |
| 31 | 65 | 25.8 | 772 | 2 T27512 | hypothetical prote |
| 32 | 65 | 25.8 | 972 | 2 S35521 | DNA topoisomerase |
| 33 | 64.5 | 25.6 | 190 | 2 A24713 | sericin - silkworm |
| 34 | 64.5 | 25.6 | 492 | 2 S63065 | hypothetical prote |
| 35 | 64.5 | 25.6 | 510 | 2 T37541 | probable glycolipl |
| 36 | 64 | 25.4 | 250 | 2 B35026 | filaggrin B - mous |
| 37 | 64 | 25.4 | 255 | 2 A35026 | filaggrin A - mous |
| 38 | 64 | 25.4 | 292 | 2 T23966 | hypothetical prote |
| 39 | 64 | 25.4 | 344 | 2 T40167 | hypothetical prote |
| 40 | 64 | 25.4 | 505 | 2 T05502 | hypothetical prote |
| 41 | 64 | 25.4 | 764 | 2 T39878 | hypothetical prote |
| 42 | 64 | 25.4 | 1788 | 2 T31095 | vitellogenin precu |
| 43 | 64 | 25.4 | 1912 | 2 T29088 | vitellogenin I pre |
| 44 | 63.5 | 25.2 | 328 | 2 T00747 | RING-H2 finger pro |
| 45 | 63.5 | 25.2 | 438 | 2 C86244 | DnaJ homolog, 4706 |

ALIGNMENTS

```

RESULT 1
A35938
profilaggrin - human (fragments)
C:Species: Homo sapiens (man)
C>Date: 14-Dec-1990 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
C:Accession: A35938
R:Gan, S.O.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinert, P.M.
Biochemistry 29, 9432-9440, 1990
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.
A:Reference number: A35938; MUID:91064347
A:Accession: A35938
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-2248 <GAN>
A:Cross-references: GB:J02929
C:Genetics:
A:Gene: GDB:FLG
A:Cross-references: GDB:119912; OMIM:135940
A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat
F:246-569/Region: filaggrin repeat
F:570-893/Region: filaggrin repeat
F:1074-1397/Region: filaggrin repeat
F:1573-1896/Region: filaggrin repeat

Query Match 100.0% Score 252; DB 2; Length 2248;
Best Local Similarity 100.0%; Pred. No. 2.6e-19;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 STGSGSQSHHTTQGRSDASRSGSRSTSRTRDQSGSGSRHSGS 49
DB 316 STGSGSQSHHTTQGRSDASRSGSRSTSRTRDQSGSGSRHSGS 364

RESULT 2
A32947
filaggrin precursor - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 20-Dec-1988 #sequence_revision 04-Sep-1992 #text_change 29-Sep-1999
C:Accession: A32947
R:McKinnley-Grant, L.J.; Idler, W.W.; Bernstein, I.A.; Parry, D.A.D.; Cannizzaro, L.;
Proc. Natl. Acad. Sci. U.S.A. 86, 4848-4852, 1989
A:Title: Characterization of a cDNA clone encoding human filaggrin and localization o
A:Reference number: A32947; MUID:89296901
A:Accession: A32947
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-416 <MCK>
A:Cross-references: GB:M24355; NID:9182604; PIDN:AAA52454.1; PID:9182605
A>Note: the authors translated the codon CAC for residue 188 as Gln, and AAT for resi

```

C:Genetics:
A:Gene: GDB:FLC
A:Cross-references: GDB:119912; OMIM:135940
A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat

Query Match 82.1%; Score 207; DB 2; Length 416;
Best Local Similarity 81.2%; Pred. No. 3.5e-15;
Matches 39; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 TGHSGSHSHHTTGGSRDASRGSSGSRSTRETROEGSGGSRHSGS 49
Db 97 SGHSGSHSHHTTGGSRDASRGSSGSRSTRETROEGSGGSRHSGS 144

RESULT 3
A:Accession: A45135
A:Title: Characterization of the human epidermal profilaggrin gene. Genomic organization

A:Species: Homo sapiens (man)
C:Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A45135
R:Presland, R.B.; Haydock, P.V.; Fleckman, P.; Nirunskiti, W.; Dale, B.A.
J. Biol. Chem. 267, 23772-23781, 1992
A:Title: Characterization of the human epidermal profilaggrin gene. Genomic organization
A:Reference number: A45135; MUID:93054736
A:Accession: A45135
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-591 <PRE>
A:Cross-references: GB:L01089; GB:M90967; NID:g190408; PIDN:AAA60177.1; PID:g553621
A:Note: sequence extracted from NCBI backbone (NCBIP:118773).
C:Genetics:

A:Gene: GDB:FLG
A:Cross-references: GDB:119912; OMIM:135940
A:Map position: 1q21-1q21
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand; epidermis; polymorphism; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 82.1%; Score 207; DB 2; Length 591;
Best Local Similarity 79.2%; Pred. No. 4.9e-15;
Matches 38; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

OY 2 TGHSGSHSHHTTGGSRDASRGSSGSRSTRETROEGSGGSRHSGS 49
Db 539 SGHSGSHSHHTTGGSRDASRGSSGSRSTRETROEGSGGSRHSGS 586

RESULT 4
T05221
hypothetical protein F17I5.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05221
R:Bevan, M.; Vitale, D.; Liguori, R.; Arifidou, A.; De Simone, V.; Hohelsel, J.; Mewes,
submitted to the Protein Sequence Database, July 1998
A:Reference number: Z15404
A:Accession: T05221
A:Molecule type: DNA
A:Residues: 1-343 <BEV>
A:Cross-references: EMBL:AL031032
C:Genetics:
A:Experimental source: cultivar Columbia; BAC clone F17I5
A:Map position: 4
A:Note: F17I5.120

Query Match 30.6%; Score 77; DB 2; Length 343;
Best Local Similarity 42.0%; Pred. No. 0.3;
Matches 21; Conservative 4; Mismatches 19; Indels 6; Gaps 2;

OY 5 SGSHSH---TTTGGSRDASRGSSGSRSTSR---ETRDQEGDGSRRHG 48
Db 89 SGSHSHSGTGTTHNGSHSSGSHSSATGSTRHNGTGTGSHSSGNGSRHNG 138

RESULT 5
S43693
penicillin-binding protein 2 - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 18-Feb-2000
C:Accession: S43693
R:Murakami, K.; Fujimura, T.; Doi, M.
FEBS Microbiol. Lett. 117, 131-136, 1994
A:Title: Nucleotide sequence of the structural gene for the penicillin-binding protein

A:Reference number: S43693; MUID:94237445
A:Accession: S43693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-716 <MUR>
A:Cross-references: EMBL:X62288; NID:g483533; PIDN:CAA44177.1; PID:g483534
C:Superfamily: penicillin-binding protein 1B

Query Match 29.2%; Score 73.5; DB 2; Length 716;
Best Local Similarity 45.9%; Pred. No. 1.5;
Matches 17; Conservative 6; Mismatches 11; Indels 3; Gaps 1;

OY 5 SGSHSH---TTTGGSRDASRGSSGSRSTRETROE 38
Db 679 SGSDNNNTNRSHGSDTSANSCTAQSNNNTRSOQ 715

RESULT 6
B47328
natural killer cell tumor-recognition protein - mouse
N:Alternate names: cyclophilin-related NK-tumor recognition protein; natural killer-t
C:Species: Mus musculus (house mouse)
C:Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 05-Nov-1999
C:Accession: B47328; I77662
R:Anderson, S.K.; Gallinger, S.; Roder, J.; Frey, J.; Young, H.A.; Ortaldo, J.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 542-546, 1993
A:Title: A cyclophilin-related protein involved in the function of natural killer cell

A:Reference number: A47328; MUID:93133824
A:Accession: B47328
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1507 <AND>
A:Cross-references: GB:L04289; NID:g192866
A:Note: authors translated the codon AGT for residue 972 as Arg

R:Rinfret, A.; Anderson, S.K.
Mol. Immunol. 30, 1307-1313, 1993
A:Title: IL-2 regulates the expression of the NK-TR gene via an alternate RNA splicing
A:Reference number: I57820; MUID:94019422
A:Accession: I77662
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 235-237/263-294 <RTN>
A:Cross-references: GB:S65998; NID:g425701; PIDN:AAB28500.1; PID:g425702
C:Genetics:

A:Gene: NK-TR
C:Superfamily: natural killer cell tumor-recognition protein; cyclophilin homology
C:Keywords: alternative splicing; lymphocyte
F:60-230/Domain: cyclophilin homology <CYF>

Query Match 29.2%; Score 73.5; DB 2; Length 1507;
Best Local Similarity 39.3%; Pred. No. 3.1;
Matches 22; Conservative 6; Mismatches 21; Indels 7; Gaps 2;

OY 1 STGHSGSHSHHTTGGSR---DASRGSSGSRSTRETROEGS---GDGSRHSGS 49
Db 740 SSGRGSSKSSRSSRSRYSTRSRSLPTSRSLSPSSRSHSPMKYSDGSHSRS 795

RESULT 7
 JC4525
 nucleic acid-binding protein E5.1 - human
 N:Alternate names: E5.1 protein
 C:Species: Homo sapiens (man)
 C:Date: 08-Feb-1996 #sequence_revision 19-Apr-1996 #text_change 24-Sep-1999
 C:Accession: JC4525
 R:Badolato, J.; Gardiner, E.; Morrison, N.; Eisman, J.
 Gene 166, 323-327, 1995
 A:Title: Identification and characterization of a novel human RNA-binding protein.
 A:Reference number: JC4525; MUID:96125212
 A:Accession: JC4525
 A:Molecule type: mRNA
 A:Residues: 1-305 <BAD>
 A:Cross-references: GB:L37366; NID:g1236282; PIDN:AAA92859.1; PID:g598231
 C:Comment: This protein functions in RNA metabolism, and is involved in processing of pre-mRNA.
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins
 C:Keywords: RNA binding
 F:69-121/Region: serine-rich
 F:128-154,269-298/Region: arginine-rich
 F:162-241/Domain: RNA binding #status predicted <RNM>
 F:162-230/Domain: ribonucleoprotein repeat homology <RRM1>
 F:163-168/Region: ribonucleoprotein-2 heptamer
 F:203-210/Region: ribonucleoprotein-1 octamer
 F:278-287/Region: nuclear location signal

Query Match 28.8%; Score 72.5; DB 2; Length 305;
 Best Local Similarity 43.1%; Pred. No. 0.82;
 Matches 22; Conservative 4; Mismatches 20; Indels 5; Gaps 2;
 QY 1 STGHSSGSHHTTQGRSDASRGS-SGSRSTSRRTDQSGDGS---RH 46
 DB 81 STSSGSSSTSTSSGSSSSSSASRSGSSSTSRSSSSSSSSGSPSPRRRH 131

RESULT 8
 T29343
 hypothetical protein F21C10.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29343
 R:Du, Z.; Gattung, S.
 Submitted to the EMBL Data Library, April 1996
 A:Description: The sequence of C. elegans cosmid F21C10.
 A:Reference number: Z20610
 A:Accession: T29343
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-196 <DU>
 A:Cross-references: EMBL:U55364; PIDN:AAA97976.1; GSPDB:GN00023; CESP:F21C10.4
 A:Experimental source: strain Bristol N2; clone F21C10
 C:Genetics:
 A:Gene: CESP:F21C10.4
 A:Map position: 5
 A:Introns: 59/2; 106/2; 178/2

Query Match 28.2%; Score 71; DB 2; Length 196;
 Best Local Similarity 47.1%; Pred. No. 0.77;
 Matches 16; Conservative 3; Mismatches 15; Indels 0; Gaps 0;
 QY 16 GRSDASRGSSGSRSTSRRTDQSGDGSRHSGS 49
 DB 27 GRGGGRCGSSGARSSSRFSRSGAGAGKRFHSS 60

RESULT 9
 T08745
 probable RNA helicase - human
 N:Alternate names: protein DKFz586A0419.1

C:Species: Homo sapiens (man)
 C:Date: 11-Jun-1999 #sequence_revision 18-Aug-2000 #text_change 19-Jan-2001
 C:Accession: T08745; T13148
 R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, March 1999
 A:Reference number: Z16471
 A:Accession: T08745
 A:Molecule type: mRNA
 A:Residues: 455-819 <WAM>
 A:Cross-references: EMBL:AL050096; PIDN:CAB43268.1
 A:Experimental source: adult uterus; clone DKFz586A0419
 R:Suk, K.; Lee, M.S.
 submitted to the EMBL Data Library, August 1998
 A:Description: A novel human cDNA homologous to RNA helicase genes.
 A:Reference number: Z17605
 A:Accession: T13148
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-698; 'PSGSRPDMQSA' <SUK>
 A:Cross-references: EMBL:AF083255; NID:g3435311; PID:g3435312; PIDN:AMC32396.1
 C:Superfamily: ATP-dependent RNA helicase DBP1
 C:Keywords: ATP; DEAD box; nucleotide binding; P-loop
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 F:284-289/Region: nucleotide-binding motif B
 F:288-291/Region: DEAD motif

Query Match 28.0%; Score 70.5; DB 2; Length 819;
 Best Local Similarity 34.1%; Pred. No. 3.6;
 Matches 15; Conservative 8; Mismatches 16; Indels 5; Gaps 1;
 QY 3 GHSGSGSHHTTQGRSDASRGSSGSRSTSRRTDQSGDGSRH 46
 DB 683 GSNKRRERYENKSSSRHSHGETGNRHS-----DSPRHGGGRH 721

RESULT 10
 T33997
 hypothetical protein W03G1.5 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33997
 R:Pauley, A.; Scheel, P.; Harper, M.
 submitted to the EMBL Data Library, February 1999
 A:Description: The sequence of C. elegans cosmid W03G1.
 A:Reference number: Z21454
 A:Accession: T33997
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-471 <PAU>
 A:Cross-references: EMBL:AF125964; PIDN:AD14753.1; GSPDB:GN00022; CESP:W03G1.5
 A:Experimental source: strain Bristol N2; clone W03G1
 C:Genetics:
 A:Gene: CESP:W03G1.5
 A:Map position: 4

Query Match 27.6%; Score 69.5; DB 2; Length 471;
 Best Local Similarity 40.4%; Pred. No. 2.7;
 Matches 21; Conservative 7; Mismatches 17; Indels 7; Gaps 2;
 QY 1 STGHSSGSHHTTQGRSDASRGSSGSRSTSRRTDQSGDGSRHSGS 49
 DB 120 SSSSSFSKVSH---GRSESSSSSGSSSDDENRRKKKGRSGSSSSSGS 167

RESULT 11
 S52714
 sericin1B - silkworm
 C:Species: Bombyx mori (silkworm)
 C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
 C:Accession: S52714
 R:Garel, A.A.; Deleage, G.G.; Prudhomme, J.J.

| | | | | |
|-----------------------|----------------|---------------|----------|------------|
| Query Match | 27.4% | Score 69 | DB 2 | Length 284 |
| Best Local Similarity | 43.2% | Pred. NO. 1.8 | | |
| Matches 16 | Conservative 6 | Mismatches 15 | Indels 0 | Gaps 0 |

Search completed: January 28, 2002, 07:38:59
Job time: 89 sec

This Page Blank (uspto)

| | |
|---|---|
| RESULT | 2 |
| ID | RTOA_DICD1 |
| AC | P54681; |
| DT | 01-OCT-1996 (Rel. 34, Created) |
| DT | 01-OCT-1996 (Rel. 34, Last sequence update) |
| DT | 01-NOV-1997 (Rel. 35, Last annotation update) |
| DE | RTOA PROTEIN (RATIO-A). |
| CN | RTOA. |
| OS | Dicystostellum discoideum (Slime mold). |
| OC | Eukaryota; Mycetozoa; Dicystostellida; Dicystostellium. |
| OX | NCBI_TaxID=44689; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=97108762; PubMed=8951083; |
| RA | Wood S.A., Ammann R.R., Brock D.A., Li L., Spann T., Gomer R.H.; |
| RT | "RtoA links initial cell choice to the cell cycle in |
| RL | Dicystostellum."; |
| RT | Development 122:3677-3685(1996). |
| CC | -I- FUNCTION: MAY HAVE DUAL FUNCTIONS, ONE IN VEGETATIVE CELLS AT STARVATION INVOLVING CELL-TYPE CHOICE AND A LATER FUNCTION INVOLVING PROGRESSION TO THE TIPPED MOUND STAGE. |
| CC | -I- DEVELOPMENTAL STAGE: EXPRESSED IN VEGETATIVE CELLS. LEVELS DECREASE DRAMATICALLY AS DEVELOPMENT BEGINS, AND REMAIN LOW THROUGHOUT THE AGGREGATION AND LOOSE MOUND STAGES. THE LEVELS BECOME MORE ABUNDANT AGAIN AT THE TIGHT-MOUND STAGE AND REMAIN HIGH THROUGHOUT THE REST OF DEVELOPMENT AND FRUITING BODY FORMATION. |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). |
| CC | ----- |
| DR | EMBL; U48298; AAC47311.1; . |
| DK | Dictydb; DD02056; rtoa. |
| KW | Transmembrane; Repeat. |
| FT | TRANSMEM 30 50 POTENTIAL. |
| FT | DOMAIN 30 38 POLY-LBU. |
| FT | DOMAIN 87 96 POLY-SER. |
| FT | DOMAIN 181 181 11 X 11 AA TANDEM REPEATS OF G-S-(2)-N-S-G-S-E-S(3). |
| FT | REPEAT 181 190 1 (INCOMPLETE). |
| FT | REPEAT 191 200 2 (INCOMPLETE). |
| FT | REPEAT 201 211 3. |
| FT | REPEAT 212 222 4. |
| FT | REPEAT 223 233 5. |
| FT | REPEAT 234 244 6. |
| FT | REPEAT 245 255 7. |
| FT | REPEAT 256 266 8. |
| FT | REPEAT 267 277 9. |
| FT | REPEAT 278 287 10 (INCOMPLETE). |
| FT | REPEAT 288 298 11. |
| SQ | SEQUENCE 400 AA; 39840 MW; 0392F4E6BDC27A75 CRC64; |
| Query Match | 31.7%; Score 80; DB 1; Length 400; |
| Best Local Similarity | 37.3%; Pred. No. 0.06; |
| Matches 19; Conservative 10; Mismatches 20; Indels 2; Gaps 1; | |
| OY | 1 STGSSNSHSHHTTCGRSDASRGSSGSRSRSTRTRDQEOGSGDSRRSGS 49 |
| Db | 179 STGSSNSGESSESGSSNSGESSESGSSNSGESSESGSSNSGS 229 |
| RESULT | 3 |
| ID | NKCR_MOUSE |
| AC | P30415; |
| DT | 01-APR-1993 (Rel. 25, Created) |

| | |
|--------------------------|--|
| DT | 30-MAY-2000 (Rel. 39, Last sequence update) |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) |
| DE | NK-TUMOR RECOGNITION PROTEIN (NATURAL-KILLER CELLS CYCLOPHILIN- |
| DE | RELATED PROTEIN) (NK-TR PROTEIN). |
| OS | Mus musculus (Mouse). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| OX | NCBI_TaxID=10090; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RX | MEDLINE=93133824; PubMed=8421688; |
| RA | Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A., |
| RA | Orraldo J.R.; |
| RT | "A cyclophilin-related protein involved in the function of natural |
| RT | killer cells."; |
| RL | proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993). |
| RM | [2] |
| RP | REVISIONS TO C-TERMINUS. |
| RC | STRAIN-BALB/C; TISSUE=Blood; |
| RA | Anderson S.K., |
| RL | Submitted (Oct-1999) to the EMBL/GenBank/DDBJ databases. |
| CC | -I- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX. |
| CC | INVOLVED IN THE FUNCTION OF NK CELLS. |
| CC | -I- SIMILARITY: CONTAINS A CYCLOPHILIN-LIKE PPIASE DOMAIN. |
| CC | ----- |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration |
| CC | between the Swiss Institute of Bioinformatics and the EMBL Outstation - |
| CC | the European Bioinformatics Institute. There are no restrictions on its |
| CC | use by non-profit institutions as long as its content is in no way |
| CC | modified and this statement is not removed. Usage by and for commercial |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ |
| CC | or send an email to license@isb-sib.ch). |
| CC | ----- |
| DR | EMBL; L04289; AAA37500.2; ALT_INIT. |
| DR | HSSP; Q27450; IAS8. |
| DR | MCD; MG1:97346; NKTf. |
| DR | InterPro: IPRO02130; CSA_PPIase. |
| DR | Pfam: PF00160; pro_isomerase; 2. |
| DR | PROSITE; PS00170; CSA_PPIASE_1; 1. |
| DR | PROSITE; PS50072; CSA_PPIASE_2; 1. |
| KW | Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane. |
| FT | DOMAIN 1 176 |
| FT | DOMAIN 222 241 ARG/LYS-RICH (BASIC). |
| FT | DOMAIN 422 459 ARG/LYS-RICH (BASIC). |
| FT | DOMAIN 964 1003 ARG/LYS-RICH (BASIC). |
| FT | DOMAIN 198 273 ARG/SER-RICH. |
| FT | DOMAIN 468 565 ARG/SER-RICH. |
| FT | DOMAIN 658 812 ARG/SER-RICH. |
| FT | DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH. |
| SQ | SEQUENCE 1453 AA; 163439 MW; DF1173FFB14B283E CRC64; |
| Query Match | 29.2%; Score 73.5; DB 1; Length 1453; |
| Best Local Similarity | 39.3%; Pred. No. 1.6; |
| Matches 22; Conservative | 39.3%; 6; Mismatches 21; Indels 7; Gaps 2; |
| QY | 1 STGHSQSCHSHTTQGRS--DASRGSSGSRSSTHETRDQDS---GDGSRHGSGS 49 |
| I: | : : : : : |
| Db | 686 SSGRGGSKSSRRSSRSRYTSRSLPTSRSLSPSSRSHSPMKYDGSGHSRS 741 |
| RESULT 4 | |
| VLPB_MYCHR | VLBP_MYCHR STANDARD; PRT; 174 AA. |
| ID | P29229; |
| DT | 01-DEC-1992 (Rel. 24, Created) |
| DT | 01-DEC-1992 (Rel. 24, Last sequence update) |
| DT | 01-NOV-1997 (Rel. 35, Last annotation update) |
| DE | VARIANT SURFACE ANTIGEN B PRECURSOR (VLBP PROLIPOPROTEIN). |
| GN | VLBP. |
| OS | Mycoplasma hyorhinis. |
| OC | Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes; |
| OC | Mycoplasmataceae; Mycoplasma. |

```

OX NCBI_TaxID=2100;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SK76;
RA MEDLINE=92097525; PubMed=1722868;
RA Yoven D., Rosengarten R., Watson-Mckown R., Wise K.S.;
RT "Molecular basis of Mycoplasma surface antigenic variation: a novel
RT set of divergent genes undergo spontaneous mutation of periodic
RT coding regions and 5' regulatory sequences."
RL EMO J. 10:4069-4079(1991).
CC -1- FUNCTION: RESPONSIBLE FOR THE ANTIGENIC DIVERSITY FOR HOST
CC ADAPTATION.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (PROBABLE).
CC -1- MISCELLANEOUS: THE NUMBERS OF REPEATS CAN VARY AND IS ONE OF THE
CC BASIS OF THE ANTIGENIC DIVERSITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62936; CAA44709.1; ALT_SEQ.
DR PIR: S18654; S18654.
DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
KW Antigen; Membrane; Lipoprotein; Repeat; Signal.
FT SIGNAL 1 29 PROBABLE.
FT CHAIN 30 174 VARIANT SURFACE ANTIGEN B.
FT LIPID 88 30 N-ACYL DIGLYCERIDE (PROBABLE).
FT DOMAIN 88 171 7 X 12 AA TANDEM REPEATS.
FT REPEAT 100 111 1.
FT REPEAT 112 123 2.
FT REPEAT 124 135 3.
FT REPEAT 136 147 4.
FT REPEAT 148 159 5.
FT REPEAT 160 171 6.
FT REPEAT 174 174 7.
SQ SEQUENCE 174 AA; 16145 MW; 28F4C9ECA85585D7 CRC64;

Query Match 27.4%; Score 69; DB 1; Length 174;
Best Local Similarity 43.1%; Pred. NO. 0.5;
Matches 22; Conservative 7; Mismatches 20; Indels 2; Gaps 2;

OY 1 STGHGSG-OHSHTTTOGRSDASRGSGSRSTSTRTDQEOGSGSRHSGS 49
Db 48 SGGGSGTGLGSGTTTGGSGGTTTGGSSSSSTTGGCGTGTGSDSDSGA 98

RESULT 5
SR40_YEAST STANDARD: PRT; 406 AA.
AC P32583;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SUPPRESSOR PROTEIN SRP40.
GN SRP40 OR YKR092C OR YKR412A.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 28583 / FL100;
RA Laio D., Carles C., Sentenac A., Thuriaux P.;
RL Submitted (May-1993) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC MEDLINE=94205265; PubMed=8154186;

```

```

RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
RA Remacha M., Jimenez A., del Rey F., Ballesta J.P.G., Revuelta J.L.;
RT "The complete sequence of a 15,820 bp segment of Saccharomyces
RT cerevisiae chromosome XI contains the UBI2 and MUP1 genes and three
RT new open reading frames."
RL Yeast 9:1349-1354(1993).
CC -1- FUNCTION: NOT KNOWN. WEAK SUPPRESSOR OF A MUTANT OF THE
CC SUBUNIT AC40 OF DNA DEPENDANT RNA POLYMERASE I AND III.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L11275; AAA35091.1; -
DR EMBL: X73541; CAA51946.1; -
DR EMBL: Z28317; CAA82171.1; -
DR PIR: S38170; S38170.
DR SCD: S0001800; SRP40.
FT DOMAIN 25 314 ASP/SER-RICH.
FT CONFLICT 400 400 G -> N (IN REF. 1).
SQ SEQUENCE 406 AA; 41015 MW; 8EA007695AF4BAID CRC64;

Query Match 27.0%; Score 68; DB 1; Length 406;
Best Local Similarity 32.7%; Pred. NO. 1.6;
Matches 16; Conservative 8; Mismatches 25; Indels 0; Gaps 0;

OY 1 STGHGSG-OHSHTTTOGRSDASRGSGSRSTSTRTDQEOGSGSRHSGS 49
Db 158 SGGSSSSSESGSDSDSSSSSSSSSSSDSDSDSDSSSSSSSSSS 206

RESULT 6
AF4_MOUSE STANDARD: PRT; 1217 AA.
AC O88573;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE AF-4 PROTEIN (PROTO-ONCOGENE AF4).
GN MLT12 OR MLT12H.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RA Isnard P.;
RL Submitted (JUN-1998) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE AF4 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF074266; AAD08668.1; -
DR MGD: MGI:1100819; MLT12h.
KW Nuclear protein; Proto-oncogene.
FT DOMAIN 261 266 POLY-PRO.
FT DOMAIN 467 476 POLY-SER.
FT DOMAIN 477 483 POLY-GLU.
FT DOMAIN 841 845 POLY-SER.
FT DOMAIN 871 881 POLY-SER.

```


| | |
|-----------------------|---|
| DR | InterPro: IPR002154; Neurugulin. |
| DR | Pfam: PF00008; EGF_1. |
| DR | Pfam: PF00047; Ig_1. |
| DR | Pfam: PF02158; Neurugulin: 2. |
| DR | SMART: SM00181; EGF_1. |
| DR | SMART: SM00408; IgC2; 1. |
| DR | PROSITE: PS00022; EGF_1; 1. |
| DR | PROSITE: PS01186; EGF_2; 1. |
| KW | Growth factor; EGF-like domain; Immunoglobulin domain; Glycoprotein; Transmembrane; Multigene family; Alternative splicing. |
| FT | PROPEP 1 127 |
| FT | CHAIN 128 868 |
| FT | CHAIN 128 428 |
| FT | DOMAIN 128 429 |
| FT | TRANSMEM 430 450 |
| FT | DOMAIN 451 868 |
| FT | DOMAIN 266 334 |
| FT | DOMAIN 346 356 |
| FT | DOMAIN 357 398 |
| FT | DOMAIN 22 32 |
| FT | DOMAIN 35 45 |
| FT | DOMAIN 56 59 |
| FT | DOMAIN 103 106 |
| FT | DOMAIN 739 745 |
| FT | DISULFID 273 327 |
| FT | DISULFID 361 375 |
| FT | DISULFID 369 386 |
| FT | DISULFID 388 397 |
| FT | CARBOHYD 33 33 |
| FT | CARBOHYD 34 34 |
| FT | CARBOHYD 163 163 |
| FT | CARBOHYD 294 294 |
| FT | CARBOHYD 362 362 |
| FT | VARSPLIC 1 108 |
| FT | VARSPLIC 220 222 |
| FT | VARSPLIC 388 388 |
| FT | VARSPLIC 389 868 |
| FT | VARSPLIC 390 412 |
| FT | VARSPLIC 390 421 |
| FT | VARSPLIC 414 421 |
| FT | VARSPLIC 414 439 |
| FT | VARSPLIC 440 868 |
| FT | CONFLICT 117 117 |
| FT | CONFLICT 724 724 |
| SQ | SEQUENCE 868 AA; 93776 MW; 3C7d4D94DBE64DE2 CMC64; |
| OY | 1 STGHSSGSHHTTQGRSDAKSGSSGSSTSRTRPDQDSGDGS 44 |
| Db | 26 SSSSSSSNNSSSSTSSRSSRSRSCSTTTTSSENSGNS 69 |
| Matches | 15; Conservative 7; Mismatches 22; Indels 0; Gaps |
| Query Match | 26.2%; Score 66; DB 1; Length 868; |
| Best Local Similarity | 34.1%; Pred. No. 5.8; |
| RESULT 11 | |
| ID | SER1_GALME STANDARD; PRT; 115 AA. |
| AC | 096614; |
| DT | 20-AUG-2001 (Rel. 40, Created) |
| DT | 20-AUG-2001 (Rel. 40, Last sequence update) |
| DT | 20-AUG-2001 (Rel. 40, Last annotation update) |
| DE | SEKICN-1 (SILK GDM PROTEIN 1) (FRAGMENT). |
| GN | SER1 OR SER-1. |
| OS | Galleria mellonella (Max moth). |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; |
| OC | Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; |
| CC | Pyralioidea; Pyralidae; Galleriinae; Galleria. |

| | |
|----|--|
| 01 | NCBI_TaxID=7137; |
| 02 | [1] |
| 03 | SEQUENCE FROM N.A. |
| 04 | TISSUE-Middle silk gland; |
| 05 | Zuvovec M., Sehnael F., Scheller K., Kumaran A.K.; |
| 06 | "silk gland specific cDNAs from Galleria mellonella L.," |
| 07 | Insect Biochem. Mol. Biol. 22:55-67(1992). |
| 08 | -1- FUNCTION: PROVIDES THE SILK FIBROIN THREAD WITH A STICKY COATING. |
| 09 | ACTS AS A CEMENT BY STRICKING SILK THREADS TOGETHER. |
| 10 | -1- SUBCELLULAR LOCATION: SECRETED. |
| 11 | -1- TISSUE SPECIFICITY: PRODUCED EXCLUSIVELY IN THE MIDDLE (MSG) |
| 12 | SECTION OF SILK GLANDS. |
| 13 | ----- |
| 14 | This SWISS-PROT entry is copyright. It is produced through a collaboration |
| 15 | between the Swiss Institute of Bioinformatics and the EMBL outstation -- |
| 16 | the European Bioinformatics Institute. There are no restrictions on its |
| 17 | use by non-profit institutions as long as its content is in no way |
| 18 | modified and this statement is not removed. Usage by and for commercial |
| 19 | entities requires a license agreement (See http://www.isb-sib.ch/announcement/ |
| 20 | or send an email to license@isb-sib.ch). |
| 21 | ----- |
| 22 | EMBL; AF095241; AAC79078.1; - |
| 23 | InterPro; IPR000087; Collagen. |
| 24 | Pfam; PF01391; Collagen; 1. |
| 25 | Silk. |
| 26 | NON TER |
| 27 | 1 1 |
| 28 | SEQUENCE 115 AA: 9161 MW: 76678FCCA95E4999 CR664; |

```

Query Match          26.0%; Score 65.5; DB 1; Length 115;
Best Local Similarity 40.8%; Pred. No. 0.77;
Matches 20; Conservative 6; Mismatches 20; Indels 3; Gaps 2;

Oy      1  STGSGSCHSHTTTQGRSDASRGSSGSSTRETRDQGSDGSRHSS 49
       1 : 1 1 1 1 1 : : : 1 1 1 1 1 : : : 1 1 1 1 1
Db      53  SSSSGS--SSSSSGSSGSSGSSGSSGSSSSSSSSSSSSSGS-SSGS 98

RESULT 12
TOP1_DROME
ID      TOP1_DROME        STANDARD;    PRT;    972 AA.
AC      P301B9; O9VXM6;
DT      01-APR-1993 (Rel. 25, Created)
DT      01-APR-1993 (Rel. 25, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      DNA TOPOISOMERASE I (Ec 5.99.1.2).
GN      TOP1 OR CG6146.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_TaxId=7227;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=93117086; Pubmed=1335568;
RA      Hsieh T.-S., Brown S.D., Huang P., Fostel J.;
RT      "Isolation and characterization of a gene encoding DNA topoisomerase
RT      I in Drosophila melanogaster."
RL      Nucleic Acids Res. 20:6177-6182(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=963564910; Pubmed=8769417;
RA      Zhang C.X., Lee M.P., Chen A.D., Brown S.D., Hsieh T.-S.;
RT      "Isolation and characterization of a Drosophila gene essential for
RT      early embryonic development and formation of cortical cleavage
RT      furrows."
RL      J. Cell Biol. 134:923-934(1996).
RN      [3]
RP      SEQUENCE FROM N.A.
RX      STRAIN=BERKELEY;
RA      Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,

```

RA Amanosides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brooksstein P., Brothier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferriz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostal M., Houston K.A., Howland T.J., Wei M.-H., Ibbegam C.,
RA Jostali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kotira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., Meled M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Slden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spleer E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svyrkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissensbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhang Y., Yeh R.N., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,
RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
CC
CC
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC
CC -1- CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC
CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC
CC -1- DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC
CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC
CC -1- NEGATIVE AND POSITIVE SUPERCOILS. WHEREAS PROKARYOTIC ENZYMES
CC
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
CC
CC -----
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
CC the European Bioinformatics Institute. There are no restrictions on its
CC
CC use by non-profit institutions as long as its content is in no way
CC
CC modified and this statement is not removed. Usage by and for commercial
CC
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: U80557; AAA28951.1; -
DR EMBL: M70064; AAC24158.1; -
DR EMBL: AE003498; AAF84440.1; -
DR PIR: S35521; S35521.
DR HSPR: P11387; 1A35.
DR FLYBase: FBgn0004924; Top1.
DR InterPro: IPR001631; topisomerse_1.
DR Pfam: PF01028; topoisomerse_1; 1.
DR PRINTS: PR00416; EOTPI5MRASE1.
DR SMART: SM00435; TOPEUC; 1.
DR PROSITE: PS00176; TOPOISOMERASE_I_EUK; 1.
KW Isomerase; Topoisomerase; DNA-binding.
FT DOMAIN 32 39
FT ACT_SITE 930 930 DNA CLEAVAGE (BY SIMILARITY).

| | | | | |
|----|----------|---------|-----------|------------------------|
| FT | CONFLICT | 40 | 40 | S -> H (IN REF. 3). |
| FT | CONFLICT | 46 | 46 | S -> SSS (IN REF. 3). |
| FT | CONFLICT | 201 | 201 | H -> O (IN REF. 3). |
| SO | SEQUENCE | 972 AA; | 11168 MM; | 3764BBDBDEFA30CD CRC64 |

| | | | | |
|-----------------------|-------|--|-----------------------|----------------------------------|
| Query Match | 25.8% | Score 65; | DB 1; | Length 972; |
| Best Local Similarity | 35.1% | Pred. No. 8.4; | | |
| Matches | 20; | Conservative | 5; | Mismatches 24; Indels 8; Gaps 3; |
| QY | 1 | STGH | SGSOAHHTTTCGRDASRSGGR | -----SRSRTRPOEQ---SGDDSRHSGS 49 |
| | | I | | |
| Db | 105 | SNSTRSGSSSHKDKDSSSKHAKSSGCHHKRRSKDKDRRDKDKRSGSSSRHAKSS | | 161 |

```

RESULT 13
ID YNM4_YEAST STANDARD; PRT; 492 AA.
AC P53919;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 54.9 KDA PROTEIN IN SPC98-TOM70 INTERGENIC REGION.
GN YNL124W OR N1888.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97245296; Pubmed=9090055;
RA De Antoni A., D'Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,
RA Pallavicini A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-13b from chromosome XIV of
RT Saccharomyces cerevisiae reveals an unusually high number of
RT overlapping open reading frames."
RL Yeast 13:261-266(1997).
RN [2]
RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96109932; Pubmed=8619318;
RA Maillet L., Bussereau F., Jaquet M.;
RT "A 43.5 kb segment of yeast chromosome XIV, which contains MFA2,
RT MRP2, CAP/SRY2, NAM9, FKBI/FPRI/BBP1, MOM22 and CPRI, predicts an
RT adenosine deaminase gene and 14 new open reading frames."
RL Yeast 11:1195-1209(1995).
CC - SIMILARITY: TO S.POMBE SPBC30D10.15.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z69382; CAA93381.1; -
DR EMBL: Z71400; CAA96005.1; -
DR SGD: S0005068; YNL124W.
KW Hypothetical protein.
KW SEQUENCE 492 AA; 54949 MW; 5A496278070F6255 CRC64;
Query Match 25.6%; Score 64.5; DB 1; Length 492;
Best Local Similarity 37.8%; Pred. No. 4.6;
Matches 17; Conservative 8; Mismatches 17; Indels 3; Gaps 2;
QY 1 STG-HSGSQHSHHTTQGRSDASRGSGSR--STSHETRDQESGD 42
   111 111 1 : : 11: 11111 1 : : :: 11
Db 50 STGLGSGSSSESESDSGSDSDSDSGSSSEEDSDANQDVGEDEGCD 94

```



```

YVD_SCHPO
ID YVD_SCHPO STANDARD; PRT; 510 AA.
AC 013692;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 53.7 KDA PROTEIN C11E3.13C IN CHROMOSOME I PRECURSOR.
GN SPAC11E3.13C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Harris D., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE GAS1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z98595; CAB1192.1; -
CC
DR KW Hypothetical protein; Glycoprotein; Signal.
FT CHAIN 1 22
FT SIGNAL 1 22
FT DOMAIN 23 510
FT CARBOHYD 423 488
FT CARBOHYD 423 488
FT CARBOHYD 216 216
FT CARBOHYD 252 252
FT CARBOHYD 318 318
FT CARBOHYD 337 337
FT CARBOHYD 397 397
FT CARBOHYD 491 491
FT SEQUENCE 510 AA; 53692 MW; 3D515CCDB5651087 CRC64;
SO

```

Query Match 25.6%; Score 64.5; DB 1; Length 510;
 Best Local Similarity 38.0%; Pred. No. 4.8;
 Matches 19; Conservative 5; Mismatches 25; Indels 1; Gaps 1;

```

OY 1 STGSSGSHHTTQGRSDASRGSS-GSRSTRERDDESGDGRHSGS 49
DB 427 TSGSSGSSASFTTASSSVSSGSSISGSSMSSTYTSASGSSAHSSGS 476

```

RESULT 15
 FILA_MOUSE STANDARD; PRT; 336 AA.
 AC P11088;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE FILAGRIN (FRAGMENT).
 GN FILG.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88058903; PUBMED=3680218;
 RA Rothnagel J.A., Mehrel T., Idler W.W., Roop D.R., Steinert P.M.;
 RT "The gene for mouse epidermal filaggrin precursor. Its partial
 RT characterization, expression, and sequence of a repeating filaggrin
 RT unit."
 RL J. Biol. Chem. 262:15643-15648(1987).

```

RN [2]
RP REVISIONS.
RA Rothnagel J.A.;
RL Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -1- PTM: FILAGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC OF 248 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03458; AAT75559.1; -
CC
DR PIR: A28444; A28444.
DR MGD: MGI:95553; Flg.
KW Phosphorylation; Polypeptide; Developmental protein.
FT NON_TER 1 1
FT SEQUENCE 336 AA; 35678 MW; 259F124D3AC0DB2D CRC64;
SO

```

Query Match 25.4%; Score 64; DB 1; Length 336;
 Best Local Similarity 34.5%; Pred. No. 3.5;
 Matches 20; Conservative 5; Mismatches 19; Indels 14; Gaps 3;

```

OY 3 GHSSGSHHTTQGRSDASRGSSG-----SNSTRERDDE---OSGDSRHSG 48
DB 225 GHSSRHQ--ADSPRYASRGSGGSGQSPASGCHSNRRKRPSPSSQSDSOVHSG 280

```

Search completed: January 28, 2002, 07:37:50
 Job time: 20 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 28, 2002, 07:37:30 ; Search time 22.14 Seconds
(without alignments)
323.728 Million cell updates/sec

Title: US-09-582-711-3
Perfect score: 252
Sequence: 1 STGHSGSQSHHTTQGRSDA.....TSRETRDQSGSGSRHSGS 49

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_protent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|--------------------|
| 1 | 252 | 100.0 | 990 | 4 | 015206 homo sapien |
| 2 | 252 | 100.0 | 1218 | 4 | 005331 homo sapien |
| 3 | 218 | 86.5 | 797 | 4 | 016824 homo sapien |
| 4 | 217 | 86.1 | 1084 | 4 | 001212 homo sapien |
| 5 | 213 | 84.5 | 798 | 4 | 09H4U3 homo sapien |
| 6 | 212 | 84.1 | 465 | 4 | 003838 homo sapien |
| 7 | 207 | 82.1 | 591 | 4 | 001720 homo sapien |
| 8 | 207 | 82.1 | 687 | 4 | 09H4U2 homo sapien |
| 9 | 197 | 78.2 | 322 | 4 | 075370 homo sapien |
| 10 | 81 | 32.1 | 727 | 2 | 053724 staphylococ |
| 11 | 81 | 32.1 | 727 | 2 | 09R744 staphylococ |
| 12 | 81 | 32.1 | 727 | 2 | 09R744 staphylococ |
| 13 | 79 | 31.3 | 300 | 5 | 09V195 leishmania |
| 14 | 77.5 | 30.8 | 397 | 5 | 09V518 drosophila |
| 15 | 77 | 30.6 | 343 | 10 | 081761 arabidopsis |
| 16 | 76 | 30.2 | 971 | 5 | 09V086 drosophila |
| 17 | 76 | 30.2 | 1913 | 5 | 09GR96 leucophaea |
| 18 | 74 | 29.4 | 1151 | 5 | 09VAY4 drosophila |
| 19 | 73.5 | 29.2 | 337 | 13 | 091708 xenopus lae |

| | | | | | |
|----|------|------|------|----|---------------------|
| 20 | 73.5 | 29.2 | 716 | 2 | 053729 staphylococ |
| 21 | 73 | 29.0 | 586 | 4 | 09U089 homo sapien |
| 22 | 73 | 29.0 | 717 | 4 | 09H307 homo sapien |
| 23 | 73 | 29.0 | 725 | 11 | 035691 mus musculus |
| 24 | 73 | 29.0 | 743 | 4 | 099738 homo sapien |
| 25 | 73 | 29.0 | 1002 | 10 | 09SM75 oryza sativ |
| 26 | 72.5 | 28.8 | 228 | 4 | 09NYG3 homo sapien |
| 27 | 72.5 | 28.8 | 305 | 4 | 09V308 homo sapien |
| 28 | 72.5 | 28.8 | 305 | 4 | 015287 homo sapien |
| 29 | 72.5 | 28.8 | 305 | 11 | 099M28 mus musculus |
| 30 | 72 | 28.6 | 989 | 5 | 09W254 drosophila |
| 31 | 71.5 | 28.4 | 448 | 5 | 09VW91 drosophila |
| 32 | 71.5 | 28.4 | 828 | 5 | 09YKR7 drosophila |
| 33 | 71 | 28.2 | 193 | 11 | 09D570 mus musculus |
| 34 | 71 | 28.2 | 196 | 5 | 019666 caenorhabdi |
| 35 | 70.5 | 28.0 | 365 | 4 | 09Y3V8 homo sapien |
| 36 | 70 | 27.8 | 460 | 5 | 094504 dictyostel |
| 37 | 70 | 27.8 | 811 | 2 | 09AD50 streptomyce |
| 38 | 70 | 27.8 | 912 | 5 | 09V3R1 drosophila |
| 39 | 69.5 | 27.6 | 200 | 4 | 003837 homo sapien |
| 40 | 68.5 | 27.6 | 322 | 5 | 023796 chironomus |
| 41 | 69.5 | 27.6 | 349 | 5 | 09XYH7 toxoplasma |
| 42 | 69.5 | 27.6 | 471 | 5 | 09UAY0 caenorhabdi |
| 43 | 69.5 | 27.6 | 703 | 6 | P79122 bos taurus |
| 44 | 69.5 | 27.6 | 1217 | 5 | 017240 bombyx mori |
| 45 | 69 | 27.4 | 284 | 6 | 099331 sus scrofa |

ALIGNMENTS

| RESULT | ID | PRELIMINARY: | PRT: | 990 AA. |
|---|---|--------------|------|---------|
| Q15206 | Q15206 | PRELIMINARY: | PRT: | 990 AA. |
| AC | Q15206 | PRELIMINARY: | PRT: | 990 AA. |
| DT | 01-NOV-1996 (TREMBLrel. 01, Created) | | | |
| DT | 01-NOV-1996 (TREMBLrel. 01, Last sequence update) | | | |
| DT | 01-JUN-2001 (TREMBLrel. 17, Last annotation update) | | | |
| DE | PROTILAGGRIN (FRAGMENT). | | | |
| GN | FLG. | | | |
| OS | Homo sapiens (Human). | | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. | | | |
| OX | NCBI_TaxID=9606; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=PLACENTA; | | | |
| RX | MEDLINE=91064347; PubMed=2248957; | | | |
| RA | Gan S.O., McBride O.W., Idler W.W., Markova N., Steinert P.M.; | | | |
| RT | "Organization, structure, and polymorphisms of the human profilaggrin | | | |
| RL | gene." | | | |
| RL | Biochemistry 29:9432-9440(1990). | | | |
| RN | [2] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | TISSUE=PLACENTA; | | | |
| RX | MEDLINE=91255199; PubMed=2043621; | | | |
| RA | Gan S.O., McBride O.W., Idler W.W., Markova N., Steinert P.M.; | | | |
| RT | "Organization, structure, and polymorphisms of the human profilaggrin | | | |
| RL | gene." | | | |
| RL | Biochemistry 30:5814-5814(1991). | | | |
| DR | EMBL: M60494; AAA63244.1; -; | | | |
| DR | InterPro: IPR003303; Filaaggrin. | | | |
| DR | PRINTS: PR00487; Filaaggrin. | | | |
| FT | NON_TER 990 | | | |
| SO | SEQUENCE 990 AA; 106452 MW; A8396F10F6A91991 CRC64; | | | |
| Query Match | | | | |
| Best Local Similarity 100.0%; Score 252; DB 4; Length 990; | | | | |
| Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0; | | | | |
| 1 STGHSGSQSHHTTQGRSDASGSGSRSTRETRDQSGSGSRHSGS 49 | | | | |
| | | | | |

Db 316 STGSHGSHSHHTTQGRSDASRGSGSRSTSRTRPDQSGDGRHSGS 364

RESULT 2

005331 PRELIMINARY: PRT: 1218 AA.

AC 005331;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE FLAGGRIN (PROFLAGGRIN) (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RM [1]

RP SEQUENCE FROM N.A.

RC TISSUE=FORESKIN;

RA MEDLINE=9109348; Pubmed=9417356;

RA Markova N.G., Marekov L.N., Chiprev C.C., Gan S.-Q., Idler W.W., Steinert P.M.;

RT "Profilaggrin is a major epidermal calcium-binding protein.";

RL Mol. Cell. Biol. 13:613-625(1993).

CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.

CC -1- PPM: FLAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF 324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHALIN GRANULES DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.

CC -1- POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FLAGGRIN REPEATS.

CC EMBL: M66943; AAA36487.1; -

DR HSP; P02638; ICP.

DR InterPro: IPR002048; EF-hand.

DR InterPro: IPR003303; Flaggrin.

DR InterPro: IPR001751; S100_Cabp.

DR Pfam: PF00036; ehand; 1.

DR Pfam: PF01023; S_100; 1.

DR PRINTS: PR00487; FLAGGRIN.

DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.

DR PROSITE: PS00303; S100_CABP; 1.

KW Phosphorylation; Polypeptide; Developmental protein; Calcium-binding; Polymorphism.

FT CA_BIND 19 32 SITE 1 (BY SIMILARITY).

FT CA_BIND 62 73 SITE 11 (BY SIMILARITY).

FT NON_TER 1218 1218

SO SEQUENCE 1218 AA; 133604 MW; EC195AD5285B19C2 CRC64;

Query Match 100.0%; Score 252; DB 4; Length 1218; Best Local Similarity 100.0%; Pred. No. 1,4e-22; Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STGSHGSHSHHTTQGRSDASRGSGSRSTSRTRPDQSGDGRHSGS 49

Db 538 STGSHGSHSHHTTQGRSDASRGSGSRSTSRTRPDQSGDGRHSGS 586

RESULT 3

016824 PRELIMINARY: PRT: 797 AA.

AC 016824;

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE FLAGGRIN PRECURSOR (FRAGMENT).

GN FLG.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RM [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=91064347; Pubmed=2248957;

RA Gan S.-Q., McBride O.W., Idler W.W., Markova N., Steinert P.M.;

RT "Organization, structure, and polymorphisms of the human profilaggrin gene.";

RL Biochemistry 29:9432-9440(1990).

DR EMBL: M60502; AAA63248.1; -

DR InterPro: IPR003303; Flaggrin.

DR PRINTS: PR00487; FLAGGRIN.

FT NON_TER 1

SO SEQUENCE 797 AA; 85176 MW; 60E6184763BDA86B CRC64;

Query Match 86.5%; Score 218; DB 4; Length 797; Best Local Similarity 85.4%; Pred. No. 1.3e-18; Matches 41; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 TGHSGSHSHHTTQGRSDASRGSGSRSTSRTRPDQSGDGRHSGS 49

Db 193 SGHSGSHSHHTTQGRSDASRGSGSRSTSRTRPDQSGDGRHSGS 240

RESULT 4

001212 PRELIMINARY: PRT: 1084 AA.

AC 001212; 003840; -

DT 01-NOV-1996 (TReMBLrel. 01, Created)

DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)

DE FLAGGRIN (FRAGMENT).

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RM [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PLACENTA;

RA MEDLINE=91064347; Pubmed=2248957;

RA Gan S.-Q., McBride W.O., Idler W.W., Markova N., Steinert P.M.;

RT "Organization, structure, and polymorphisms of the human profilaggrin gene.";

RL Biochemistry 29:9432-9440(1990).

CC -1- FUNCTION: FLAGGRIN AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.

CC -1- MISCELLANEOUS: FLAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF 317 AA, WHICH ARE SEPARATED BY A SHORT LINKER SEQUENCE (PROBABLY FLYOVSST). THE PRECURSOR IS DEPOSITED AS KERATOHALIN GRANULES. BY MEANS OF DEPHOSPHORYLATION AND PROTEOLYTIC CLEAVAGE FLAGGRIN IS FORMED.

CC EMBL: M60503; AAA63243.1; -

DR EMBL: M60501; AAA63243.1; JOINED.

DR InterPro: IPR003303; Flaggrin.

DR PRINTS: PR00487; FLAGGRIN.

KW Phosphorylation; Polypeptide; Developmental protein; Keratin; Intermediate filament.

FT NON_TER 1

SO SEQUENCE 1084 AA; 115271 MW; 80C4640B8D5A362D CRC64;

Query Match 86.1%; Score 217; DB 4; Length 1084; Best Local Similarity 83.7%; Pred. No. 2.2e-18; Matches 41; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 STGSHGSHSHHTTQGRSDASRGSGSRSTSRTRPDQSGDGRHSGS 49

Db 153 STGSHGSHSHHTTQGRSDASRGSGSRSTSRTRPDQSGDGRHSGS 201

RESULT 5

```

09HAU3
ID 09HAU3 PRELIMINARY: PRT: 798 AA.
AC 09HAU3.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE DJ14N1.1.2 (PROFILAGGRIN 3' END) (FRAGMENT).
CN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Latid G.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL356504; CAC3171.1; -
DR InterPro; IPR003303; Filaggrin.
DR PRINTS; PR00487; FILAGGRIN.
FT NON TER 1 1
SO SEQUENCE 798 AA; 84773 MW; F923DDA8D1290805 CRC64;

Query Match 84.5%; Score 213; DB 4; Length 798;
Best Local Similarity 83.3%; Pred. No. 5.2e-18;
Matches 40; Conservative 4; Mismatches 4; Indels 0; Gaps 0

0Y 2 TGGSHGSHHTTTCGRSDASRSGSGSRSTSRRTDQSGSDGSRHSGS 49
:||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 518 SCHSSSHSHHTTTCGRSDASRSGSGSRSTSRRTDQSGSDGSRHSGS 565

RESULT 6
ID 003838 PRELIMINARY: PRT: 465 AA.
AC 003838.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FILAGGRIN (PROFILAGGRIN) (FRAGMENT).
GN FLG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=PLACENTA;
RC MEDLINE=91064347; PubMed=2248957;
RX Gan S.-O., McBride O.W., Idler W.W., Markova N., Steinhert P.M.;
RA "Organization, structure, and polymorphisms of the human profilaggrin
RT gene."
RL Biochemistry 29:9432-9440(1990).
RN [2]
RP REVISIONS.
RX MEDLINE=91255199; PubMed=2043621;
RA Gan S.-O., McBride O.W., Idler W.W., Markova N., Steinhert P.M.;
RA "Organization, structure, and polymorphisms of the human profilaggrin
RT gene."
RL Biochemistry 30:5814-5814(1991).
RN [3]
RP FUNCTION: FILAGGRIN AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND
CC PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE
CC FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -1- POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND
CC WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FILAGGRIN
CC REPEATS.
CC -1- MISCELLANEOUS: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE,
CC INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM
CC COPIES OF 324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHYALIN
CC GRANULES. DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED &
CC PROTEOLYTICALLY CLEAVED.
CC EMBL; M60499; AAA63246.1; -.
DR InterPro; IPR003303; Filaggrin.

```

| | |
|--|---|
| DR | PRINTS: PR00487: FILAGRIN. |
| FT | NON_TER 1 1 |
| FT | NON_TER 465 465 |
| SQ | SEQUENCE 465 AA; 50280 MW; C883744C5E134097 CRC64; |
| Query Match 84.1%; Score 212; DB 4; Length 465; | |
| Best Local Similarity 81.2%; Pred. No. 4,1e-18; | |
| Matches 39; Conservative 6; Mismatches 3; Indels 0; Gaps | |
| OY | 2 TGHSGSHHTTGGSDASRGSSRSRTSETRDQSGDGSRHSS 49 |
| Db | 317 TGHSGSHHTTGGSDASRGSSRSRTSETRDQSGDGSRHSS 364 |
| RESULT | 7 |
| ID | 001720 |
| AC | 001720 PRELIMINARY; PRT; 591 AA. |
| DT | 01-NOV-1996 (TREMblrel. 01, Created) |
| DT | 01-NOV-1996 (TREMblrel. 01, Last sequence update) |
| DT | 01-JUN-2001 (TREMblrel. 17, Last annotation update) |
| DE | FILAGRIN PRECURSOR (PROFILAGRIN) (FRAGMENT). |
| GN | FIG. |
| OS | Homo sapiens (Human). |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |
| OC | Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. |
| OX | NCBI_TaxID=9606; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | TISSUE=PLACENTA; |
| RX | MEDLINE=93054736; Pubmed=1429717; |
| RA | Presland R.B., Haydock P.V., Fleckman P., Nirunsksiiri W., Dale B.A., |
| RT | "Characterization of the human epidermal profilaggrin gene. Genomic |
| RT | organization and identification of an S-100-like calcium binding |
| RL | domain at the amino terminus."; |
| RL | J. Biol. Chem. 267:23772-23781(1992). |
| CC | -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES |
| CC | DISULFID-BOND FORMATION AMONST THE INTERMEDIATE FILAMENTS DURING |
| CC | TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS. |
| CC | -1- PFM: FILAGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE, |
| CC | HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF |
| CC | 324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHALIN GRANULES. |
| CC | DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED & |
| CC | PROTEOLYTICALLY CLEAVED. |
| CC | -1- POLYMORPHISM: A NUMBER OF PROFILAGRIN ISOFORMS HAVE BEEN FOUND |
| CC | WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FILAGRIN |
| CC | REPEATS. |
| DR | EMBL; L01089; AAA60177.1; -; |
| DR | EMBL; L01090; AAA60176.1; -; |
| DR | HSSP; P02638; ICFP. |
| DR | MIM; 135940; -; |
| DR | InterPro; IPR002048; EF-hand. |
| DR | InterPro; IPR003303; Filaggrin. |
| DR | InterPro; IPR001751; S100_Cabp. |
| DR | Pfam; PF00036; efhand; 1. |
| DR | Pfam; PF01023; S_100; 1. |
| DR | PRINTS: PR00487: FILAGRIN. |
| DR | PROSITE; PS00018; EF_HAND; UNKNOWN_1. |
| DR | PROSITE; PS00303; S100_CABP; 1. |
| RW | Polymorphism. |
| FT | PROPEP 1 293 |
| FT | CHAIN 294 467 |
| FT | PROPEP 468 474 |
| FT | CHAIN 475 >591 |
| FT | CA_BIND 19 32 |
| FT | CA_BIND 62 73 |
| FT | NON_TER 591 591 |
| SQ | SEQUENCE 591 AA; 66366 MW; 381491625C75E369 CRC64; |
| Query Match 82.1%; Score 207; DB 4; Length 591; | |
| Best Local Similarity 79.2%; Pred. No. 2,1e-17; | |

